

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 11:39:05 ; Search time 72 Seconds
(without alignments)
477.481 Million cell updates/sec

Title: US-09-954-846-2
Perfect score: 1359
Sequence: 1 MAVLAPLALVSVPLRSRW.....EQVASTPTVSDGENKDK 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1359	100.0	258	21 AAY94893	Human protein clon
2	1359	100.0	272	21 AAB57113	Human prostate can
3	1326	97.6	296	20 AA113380	Amino acid sequenc
4	1326	97.6	296	22 AAU29039	Human PRO polypt
5	1326	97.6	296	22 AAB80248	Human PRO270 prote
6	1326	97.6	329	22 AAB12004	Human CGI-31 prote
7	1322	97.3	296	19 AAW75204	Human secreted pro
8	1313	96.6	293	19 AAW75226	Human secreted pro
9	1308	96.2	296	23 AAB06300	Prostaglandin E2 (
10	1259	92.6	372	22 AAB88335	Human membrane or

11	1147.5	84.4	344	22	ABG18889	Novel human diagno
12	994	73.1	255	21	AAY64913	Human 5' EST relat
13	916	67.4	216	19	AAW75237	Fragment of human
14	536	39.4	111	21	AGG01002	Human secreted pro
15	514	37.8	144	20	AAI12372	Human 5' EST secre
16	476.5	35.1	271	22	ABB65389	Drosophila melanog
17	447	32.9	131	20	AAI12369	Human 5' EST secre
18	431.5	31.8	103	22	ABG18875	Novel human diagno
19	430	31.6	148	20	AAI12370	Human 5' EST secre
20	422	31.1	98	22	ABG18866	Novel human diagno
21	422	31.1	105	21	AAY65404	Human 5' EST relat
22	336	24.7	462	22	ABG18886	Novel human diagno
23	272	20.0	82	21	AAY65402	Human 5' EST relat
24	124.5	9.2	227	21	AAB34845	Gene 45 human secr
25	124	9.1	104	23	AGG80708	S. cerevisiae TRX2
26	122.5	9.0	117	23	ABF60749	Borrelia burgdorfe
27	120.5	8.9	105	21	AAY95402	Archaeoglobus fulg
28	120.5	8.9	105	23	AAE18729	Archaeoglobus fulg
29	120.5	8.9	134	21	AAY95401	Archaeoglobus fulg
30	120.5	8.9	134	23	AAE18728	Archaeoglobus fulg
31	119	8.8	454	23	AAE23035	Human thioredoxin,
32	119	8.8	454	23	ABB97441	Novel human protei
33	117.5	8.6	115	23	ABP60799	Caenorhabditis ele
34	116.5	8.6	265	18	AAW26620	Thioredoxin-Trepon
35	116	8.5	127	23	ABP60782	Neurospora crassa
36	115	8.5	612	22	ABG93322	Human protein sequ
37	114	8.4	102	23	ABP60779	Mycoplasma pneumon
38	114	8.4	147	22	AAY55797	Propionibacterium
39	113	8.3	106	23	ABP60764	Coprinus comatus t
40	113	8.3	228	22	AAU87110	Novel central nerv
41	113	8.3	228	22	AAU23151	Novel human enzyme
42	111.5	8.2	228	22	AAU23151	Novel human enzyme
43	111.5	8.2	525	21	AGG25258	Arabidopsis thalia
44	111.5	8.2	526	21	AGG45726	Arabidopsis thalia
45	111.5	8.2	598	21	AGG45725	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA94893
ID AAY94893 standard; Protein; 258 AA.
AC AAY94893;
XX
XX
DT 12-JUN-2000 (first entry)
XX
DE Human protein clone HPL0392.

Human protein; hydrophobic domain; nutritional source; haematopoiesis; cytokine production; cell proliferation; cell differentiation; immune deficiency; infectious disease; autoimmune disorder; asthma; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; allergic reaction; osteoporosis; osteoarthritis; periodontal disease; nervous system disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury; systemic cytokine damage; tissue differentiation; contraceptive; stroke; coagulation disorder; myocardial infarction; inflammatory condition; septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour; nephritis; therapy.

XX Homo sapiens.

OS WO200005367-A2.

PN 03-FEB-2000.

PD 22-JUL-1999;

PF 99WO-JP03929.

XX 24-JUL-1998;

PR 98JP-0208820.

PR 07-AUG-1998;

PR 98JP-0224105.

PR 25-AUG-1998;

PR 98JP-0238116.

Qy	241	PVASTPTTVSDGENKKDK	258
Db	241	PVASTPTTVSDGENKKDK	258
RESULT 2			
AAB57113			
ID	AAB57113	standard; Protein; 272 AA.	
XX	AC	AAB57113;	
XX	DT	13-MAR-2001 (first entry)	
XX	DE	Human prostate cancer antigen protein sequence SEQ ID NO:1691.	
XX	KW	Human; prostate cancer; prostate cancer antigen; detection; diagnosis;	
XX	KW	neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;	
XX	KW	vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;	
XX	KW	antibacterial; gene therapy; neural; immune; reproductive; renal;	
XX	KW	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;	
XX	KW	wound; infectious disease.	
XX	OS	Homo sapiens.	
XX	PN	WO200055174-A1.	
XX	PD	21-SEP-2000.	
XX	PF	08-MAR-2000; 2000WO-US05988.	
XX	PR	12-MAR-1999; 99US-0124270.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PA	(ROSE/) ROSEN C A.	
XX	PI	Rosen CA, Ruben SM;	
XX	XX	WPI; 2000-587513/55.	
DR	N-PSDB;	AAF16316.	
XX	XX	Prostate cancer associated gene sequences, referred to as prostate	
PT	PT	cancer antigens, useful for treatment, prevention, and diagnosis of	
PT	PT	disorders such as prostate cancer -	
XX	XX	Claim 11; Page 2161-2162; 2338pp; English.	
PS	XX	AAF15566 to AAF16505 encode the human prostate cancer associated	
CC	CC	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.	
CC	CC	The prostate cancer antigens can have neuroprotective, cytostatic,	
CC	CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,	
CC	CC	nephrotropic, antiinfective, gynaecological and antibacterial activities,	
CC	CC	and can be used in gene therapy. The prostate cancer antigen	
CC	CC	polynucleotides may be used for detection of prostate cancer, chromosome	
CC	CC	identification, as chromosome markers, and for numerous other diagnostic	
CC	CC	or research purposes. The prostate cancer antigens may be used to treat	
CC	CC	disorders such as neural, immune, muscular, reproductive,	
CC	CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative	
CC	CC	disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to	
CC	CC	AAB57303 represent sequences used in the exemplification of the present	
CC	CC	invention.	
XX	XX	Sequence 272 AA;	
XX	XX	Query Match . 100.0%; Score 1359; DB 21; Length 272;	
XX	XX	Best Local Similarity 100.0%; Pred. No. 5.9e-140;	
XX	XX	Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNPCD	60
Db	15	MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNPCD	74
Qy	61	FDWREVEILMFLSAIVMMKNRRSMFLMTCKPPLYMGPEYIKYFNDKTIDDEELFRDKRVTW	120

Db 75 FDWEVEILMFLSAIVMMKNRRSMFLMTCKPPLYMGPEIKYFNDKTIDEELDRKRVTW 134
QY 121 IVEFFANNSDCOSFAPYIADLSLKYNCCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLP 180
Db 135 IVEFFANNSDCOSFAPYIADLSLKYNCCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLP 194
QY 181 TLILFQGGKEAMRRQIDKKGRAVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQ 240
Db 195 TLILFQGGKEAMRRQIDKKGRAVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQ 254
QY 241 PVASTPTTVSDGENKKDK 258
Db 255 PVASTPTTVSDGENKKDK 272
RESULT 3
Y13380
AAV13380 standard; Protein; 296 AA.
AC AAY13380;
XX
DT 25-JUN-1999 (first entry)
DE
XX
DE
XX
XX
KW Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW anti-thrombotic; wound healing; tissue repair.
XX
OS Homo sapiens.
XX
XX
PN WO9914328-A2.
XX
PD 25-MAR-1999.
XX
PF 16-SEP-1998; 98WO-US19330.
XX
PR 25-NOV-1997; 97US-0066840.
PR 17-SEP-1997; 97US-0059113.
PR 17-SEP-1997; 97US-0059115.
PR 17-SEP-1997; 97US-0059117.
PR 17-SEP-1997; 97US-0059119.
PR 17-SEP-1997; 97US-0059121.
PR 17-SEP-1997; 97US-0059122.
PR 17-SEP-1997; 97US-0059184.
PR 18-SEP-1997; 97US-0059263.
PR 18-SEP-1997; 97US-0059266.
PR 15-OCT-1997; 97US-0062125.
PR 17-OCT-1997; 97US-0062285.
PR 17-OCT-1997; 97US-0062287.
PR 21-OCT-1997; 97US-0063486.
PR 24-OCT-1997; 97US-0062814.
PR 24-OCT-1997; 97US-0062816.
PR 24-OCT-1997; 97US-0063045.
PR 24-OCT-1997; 97US-0063120.
PR 24-OCT-1997; 97US-0063121.
PR 24-OCT-1997; 97US-0063127.
PR 24-OCT-1997; 97US-0063128.
PR 27-OCT-1997; 97US-0063329.
PR 27-OCT-1997; 97US-0063327.
PR 28-OCT-1997; 97US-0063541.
PR 28-OCT-1997; 97US-0063542.
PR 28-OCT-1997; 97US-0063544.
PR 28-OCT-1997; 97US-0063549.
PR 28-OCT-1997; 97US-0063550.
PR 28-OCT-1997; 97US-0063564.
PR 29-OCT-1997; 97US-0063435.
PR 29-OCT-1997; 97US-0063704.
PR 29-OCT-1997; 97US-0063732.

PR 29-OCT-1997; 97US-0063738.
PR 29-OCT-1997; 97US-0063734.
PR 29-OCT-1997; 97US-0064215.
PR 29-OCT-1997; 97US-0063735.
PR 31-OCT-1997; 97US-0063870.
PR 31-OCT-1997; 97US-0064103.
PR 03-NOV-1997; 97US-0064248.
PR 07-NOV-1997; 97US-0064809.
PR 12-NOV-1997; 97US-0065186.
PR 17-NOV-1997; 97US-0065846.
PR 18-NOV-1997; 97US-0065693.
PR 21-NOV-1997; 97US-0066120.
PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066772.
PR 24-NOV-1997; 97US-0066466.
PR 24-NOV-1997; 97US-0066770.
PR 24-NOV-1997; 97US-0066511.
PR 24-NOV-1997; 97US-00666453.
XX
XX (GETH) GENENTECH INC.
XX
XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XX
XX WPI; 1999-229533/19.
XX N-PSDB; AAX52251.
XX New isolated human genes and polypeptides used in, e.g. treatment of
XX gastrointestinal ulceration
XX
XX Claim 12; Fig 76; 320pp; English.
XX
XX AAY13344-403 represent secreted and transmembrane human proteins.
XX The cDNA sequences are obtained from cDNA libraries, prepared from
XX fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
XX The encoded polypeptides have specific uses based on their homology to
XX known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
XX associated with the preservation and maintenance of gastrointestinal
XX mucosa and the repair of acute and chronic mucosal lesions
XX (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
XX ulceration and congenital microvillus atrophy), skin diseases associated
XX with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
XX cancers such as lung squamous cell carcinoma of the vulva and gliomas),
XX potent effects on cell growth and development, diseases related to
XX growth or survival of nerve cells including Parkinson's disease,
XX Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
XX for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
XX as a target for anti-tumor drugs. PRO533 may be used in the treatment
XX of Usher Syndrome or Atrophia areata; PRO269 can be used as an
XX anti-thrombotic agent; PRO287 polypeptides and portions may have
XX therapeutic applications in wound healing and tissue repair; PRO317 can
XX be used for treating problems of the kidney, uterus, endometrium, blood
XX vessels, or related tissue, e.g. in the heart of genital tract.
XX
SQ Sequence 296 AA;
Query Match 97.68; Score 1326; DB 20; Length 296;
Best Local Similarity 86.8%; Pred. No. 2.7e-136;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;
QY 1 MAVLAPLIAIVSYVPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 60
Db 1 MAVLAPLIAIVSYVPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 60
QY 61 FDWEVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWEVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMCKPPLYMGPEIKYFNDKTIDEELDRKRVTWIVEFFANNSDCOSFAPYIADL 142
Db 121 IVFLMCKPPLYMGPEIKYFNDKTIDEELDRKRVTWIVEFFANNSDCOSFAPYIADL 180
QY 143 SLKYNCCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLPTLILFQGGKEAMRRPOIDKKGR 202

```
Db 181 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFOGKEMRRPQIDKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
    |||||||
Db 241 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 4
AAU29039
ID AAU29039 standard; Protein; 296 AA.
AC AAU29039;
XX
XX
DT 18-DEC-2001 (first entry)
XX
XX
DE Human PRO polypeptide sequence #16.
XX
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
XX Homo sapiens.
    WO200168848-A2.
XX
PD 20-SEP-2001.
XX
XX 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196659P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
```

```
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-602746/68.
XX N-PSDB; AAS45940.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours; in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 11; Fig 32; 774pp; English.
XX
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
XX The PRO polypeptides and their associated nucleic acids can be used to
XX detect the presence of a tumour in a mammal by comparing the level of
XX expression of a PRO polypeptide in a test sample of cells from the animal
XX and a control sample of normal cells, whereby a higher level of
XX expression in the test sample indicates the presence of a tumour in the
XX mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
XX and rabbits but are preferably human. The polypeptides can be used to
XX stimulate tumour necrosis factor (TNF). Alpha release from human blood,
XX when contacted with it. A specific polypeptide can be used to stimulate
XX the proliferation or differentiation of chondrocyte cells. The PRO
XX proteins can be used to determine the presence of tumours and also
XX susceptibility to tumour development, particularly adrenal, lung, colon,
XX breast, prostate, rectal, cervical, or liver tumours, in mammalian
XX subjects. The oligonucleotide probes specific for the PRO nucleic acids
XX can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 296 AA;
QY Query Match 97.6%; Score 1326; DB 22; Length 296;
    Best Local Similarity 86.8%; Pred. NO. 2.7e-136;
    Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;
QY 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKPLPLCHGLPTQREDGNPCD 60
    |||||||
Db 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKPLPLCHGLPTQREDGNPCD 60
QY 61 FDMREVEILMFLSAIVMKNNRS----- 83
    |||||||
Db 61 FDMREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLATCKPPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANNSNDCQSFAPYADL 142
    :|||||
Db 121 IVFLATCKPPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANNSNDCQSFAPYADL 180
QY 143 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFOGKEMRRPQIDKGR 202
    |||||||
Db 181 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFOGKEMRRPQIDKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
    |||||||
Db 241 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 5
AAB80248
ID AAB80248 standard; Protein; 296 AA.
XX
XX AAB80248;
XX
XX 24-APR-2001 (first entry)
XX
XX Human PRO270 protein.
XX
XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiant;
KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW antiarthritic; antifertility; antidiabetic; antiviral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW ischaemia; inflammation.
XX
XX Homo sapiens.
```


XX WO2001043111-AL.
 XX 18-JAN-2001.
 XX 22-FEB-2000; 2000WO-US04414.
 XX 07-JUL-1999; 99US-0143048.
 XX 26-JUL-1999; 99US-0145698.
 XX 28-JUL-1999; 99US-0146222.
 XX 08-SEP-1999; 99WO-US20594.
 XX 13-SEP-1999; 99WO-US20944.
 XX 15-SEP-1999; 99WO-US21090.
 XX 15-SEP-1999; 99WO-US21547.
 XX 05-OCT-1999; 99WO-US22089.
 XX 29-NOV-1999; 99WO-US28214.
 XX 30-NOV-1999; 99WO-US28313.
 XX 16-DEC-1999; 99WO-US30095.
 XX 20-DEC-1999; 99WO-US30911.
 XX 20-DEC-1999; 99WO-US30999.
 XX 05-JAN-2000; 99WO-US00219.
 XX (GETH) GENENTECH INC.
 XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 XX Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 XX Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin LJ;
 XX Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 XX Williams PM, Wood WT;
 XX WPI: 2001-081051/09.
 XX N-PSDB; AAF72409.
 XX Sixty one nucleic acids encoding PRO polypeptides which are useful in
 XX the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 XX squamous cell carcinoma) and neurodegenerative diseases (e.g.
 XX Alzheimer's disease) -
 XX Claim 1; Fig 76; 393pp; English.
 XX The present sequence is one of sixty one novel secreted and
 XX transmembrane PRO polypeptides. The PRO polypeptides are
 XX useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 XX squamous cell carcinoma), gastrointestinal disorders (e.g.
 XX enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 XX Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 XX endometrial bleeding angiogenesis, ischaemias such as coronary
 XX ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 XX rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 XX diabetes and retinal disorders such as retinitis pigmentosum.
 XX The PRO nucleic acids have applications in molecular biology, including
 XX use as hybridization probes, and in chromosome and gene mapping.
 XX Sequence 296 AA;
 XX
 XX Query Match 97.6%; Score 1326; DB 22; Length 296;
 XX Best Local Similarity 86.8%; Pred. No. 2.7e-136;
 XX Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;
 XX
 XX 1 MAVLAPLIALVSVPRSLRWLAQPYLLSALLSAFLLVRLKPLCHGLTQREDGNPCD 60
 XX |
 XX 1 MAVLAPLIALVSVPRSLRWLAQPYLLSALLSAFLLVRLKPLCHGLTQREDGNPCD 60
 XX |
 XX 61 FDWREVEILMFLSAIYVMKNRRS-----83
 XX |
 XX 61 FDWREVEILMFLSAIYVMKNRRSITVEQHTGNFMFSKVANTILFFRLDMGLLYITLC 120
 XX |
 XX 84 -NFLMTCKPLYMGPEYIKYFNDKTIDEELEDRKRVTWIWEFFANWNCQSFAPYADL 142
 XX |
 XX 121 IVFLMTCKPLYMGPEYIKYFNDKTIDEELEDRKRVTWIWEFFANWNCQSFAPYADL 180
 XX |
 XX 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKVTSPITKQLPTLILFGGKEAMRRPOIDKKGR 202
 XX |

Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKVTSPITKQLPTLILFGGKEAMRRPOIDKKGR 240
 Qy 203 AVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQVASTPTTVSDGKNKDK 258
 Db 241 AVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQVASTPTTVSDGKNKDK 296
 RESULT 6
 ABB12004
 ID ABB12004 standard; peptide; 329 AA.
 XX AC
 XX ABB12004;
 XX
 XX 11-JAN-2002 (first entry)
 XX
 XX Human CGI-31 protein homologue, SEQ ID NO:2374.
 XX
 XX Human: cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200157188-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US03800.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 XX 27-APR-2000; 2000US-0560875.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX WPI: 2001-457740/49.
 XX N-PSDB; ABA09248.
 XX
 XX Human proteins and DNA encoding sequences useful for preventing,
 XX treating or ameliorating a medical condition in a mammalian subject
 XX e.g. arthritis and cancer -
 XX
 XX Claim 20; Page 294-295; 1963pp; English.
 XX
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 XX invention also relates to vectors and recombinant host cells comprising a
 XX nucleotide of the invention, methods of producing the novel polypeptides,
 XX antibodies against the polypeptides, methods of detecting the nucleotides
 XX or polypeptides in a sample, and methods of identifying compounds which
 XX bind to polypeptides of the invention. Although novel, many of the
 XX polypeptides of the invention have homology to known proteins, and hence
 XX giving an insight into their probable biological activities, and hence
 XX potential therapeutic applications. The polypeptides of the invention may
 XX have various activities, including cytokine, cell proliferation or cell
 XX differentiation activities; stem cell growth factor activity;
 XX haematopoiesis regulatory activity; tissue growth activity;
 XX immunomodulatory activity; activin- or inhibin-related activities;
 XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
 XX thrombolytic activities; receptor or ligand activities; or may be
 XX involved in oncogenesis, cancer cell proliferation or metastasis.
 XX depending on their biological activities, polypeptides and nucleotides of

AAW75226	
ID	AAW75226 standard; Protein: 293 AA.
XX	
XX	
AC	AAW75226;
XX	
DT	29-JAN-1999 (first entry)
XX	
DE	Human secreted protein encoded by gene 9 clone HJBCD89.
XX	
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
CC	
XX	Homo sapiens.
PN	WO9840483-A2.
XX	
PD	17-SEP-1998.
XX	
PF	12-MAR-1998; 98WO-US04858.
XX	
PR	19-DEC-1997; 97US-0068368.
PR	14-MAR-1997; 97US-0040710.
PR	14-MAR-1997; 97US-0040762.
PR	30-MAY-1997; 97US-0048100.
PR	30-MAY-1997; 97US-0048189.
PR	30-MAY-1997; 97US-0048357.
PR	30-MAY-1997; 97US-0050934.
PR	06-JUN-1997; 97US-0048970.
PR	05-SEP-1997; 97US-0057765.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
PI	Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
PI	Wei YF, Young PE, Zeng Z;
XX	
DR	WPT; 1998-520811/44.
DR	N-PSDB: AAV34316.

Isolated human polynucleotide(s) encoding secretory peptide(s) -
used to develop products for the diagnosis and treatment of e.g.
inflammation, cancers, CNS disorders or immune system disorders

Claim 1: Page 173-174; 201pp; English.

This sequence represents a secreted human protein encoded by the gene clone detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AAV34277) for increasing the stability of the fused protein as compared to the human protein only.

The invention relates to 28 novel genes and their fragments (nucleic acid sequences: AAV34286-V34325; amino acid sequences AAV75196-W75235) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy.

Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 28 polynucleotides, based on which tissues they are most highly expressed in (see AAV34286 for described uses).

XX	Sequence	293 AA:
QQ	Query Match	96.6%
	Best Local Similarity	86.7%
	Matched	1; Mismatches 0;
	Indels	38; Gaps 1;
OV	4 LAPLIAIVSYVPRLSRWLAOPYLLLSALLSAFLIVRLKLPCLCHGLPTQREDGNPCDFDW	63

Db	1	LAPLIALVYVPRLSRWLAQYVYLLSALLSNAFLVVRKLPPLCHGLPTQREDNPCDFDW	60
QY	64	REVEILMFLSAIVMMKNRRS-----MF	85
Db	61	REVEILMFLSAIVMMKNRRSITVBQHIGNIEMFSKVANTILFFRLDIRMGLYITLCIVF	120
QY	86	LMTCCKPLYMGPEYIKYFNDKTIIDBELERDKRVTWIVEFFANWSNDCOSFAPYIADLSLK	145
Db	121	LMTCCKPLYMGPEYIKYFNDKTIIDBELERDKRVTWIVEFFANWSNDCOSFAPYIADLSLK	180
QY	146	YNTCTGLNFGKVDVGRYTDVSTRYKVYSTPLKQPLTLILFOGGKAMRRPQIDDKGRAVS	205
Db	181	YNTCTGLNFGKVDVGRYTDVSTRYKVYSTPLKQPLTLILFOGGKAMRRPQIDDKGRAVS	240
QY	206	WTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGNNKDK	258
Db	241	WTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGNNKDK	293

RESULT 9

RESD11	ABB06300	standard; Protein; 296 AA.
ID	ABB06300	
XX	AC	
XX	ABB06300;	
XX		
DT	01-JUN-2002	(first entry)
XX		
DE	Prostaglandin E2 (PGE2)	protein.
XX		
KW	Prostaglandin E2; PGE2;	stomach cancer.
XX		
OS	Unidentified.	
XX		
PN	KR2001081233-A.	
XX		
PD	29-AUG-2001.	
XX		
PF	11-FEB-2000; 2000KR-0006416.	
XX		
PR	11-FEB-2000; 2000KR-0006416.	
XX		
PA	(NAHE-) NAT INST HEALTH IN KOREA.	
XX		
PI	Kim GC, Nam MJ, Park MS;	
XX		
WPI	2002-136917/18.	
DR	N-PSDB: ABL49614.	
DR		

XX Novel gene expressed by prostaglandin E2 (PGE2) in stomach cancer cell
PT
PT line snu-1 and its amino acid sequence -
XX
XX
PS Disclosure; Page 12; 12pp; Korean.
XX
XX The present sequence represents prostaglandin E2 (PGE2) which is
CC
CC expressed in the stomach cancer cell line SNU-1. The PGE2 gene can be
CC used in the treatment of stomach cancer.
XX
XX Sequence 296 AA:
SO

	Query Match	96.2%;	Score 1308;	DB 23;	Length 296;
	Best Local Similarity	86.1%;	Pred. No. 2.6e-134;		
	Matches 255;	Conservative 1;	Mismatches 2;	Indels 38;	Gaps 1
Qy	1	MAVLAPLIALYVS	PRLSRWLTAQPYYLLSALLSAFLLVKRLPPLCHGLPTQREDGNPCD	60	
Db	1	MAVLAPLIALYVS	PRLSRWLTAQPYYLLSALLSAFLLVKRLPPLCHGLPTQREDGNPCD	60	
Qy	61	FDWEVEITLMFSLAIVMMKNRRS	-----	-----	83
Db	61	FDWEVEITLMFSLAIVMMKNRRS	ITVEQHIGNIFMFSKVANTILFRDLIRMLLYITLC	120	
Qy	84	-MFMTCKPPYVMGPEYIKYFNDKTIDBELEDRDKRTWIVEEFANNSDCQSFAPIAYDL	142		

Db 121 IIVFLMTCKPPLYMGPEYIKYFNDKTIIDEELERKRVTWIIEFFANWSDCQSFAPYADL 180
 QY 143 SLKYNTCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLPTLILFQGGKEAMRRPQIDKKGR 202
 Db 181 SLKYNTCTGLNFGKVDVGRYTDVSTRYKASTSPLTKQLPTLILFQGGKEAMRRPQIDKKGR 240
 QY 203 AVSWTFSEENVIREFNLNELYQRAKKLSKAGDNIPPEQPVASTPTTYSNGENKKDK 258
 Db 241 AVSWTFSEENVIREFNLNELYQRAKKLSKAGDNIPPEQPVASTPTTYSNGENKKDK 296

RESULT 10

AAB88335
 ID AAB88335 standard; Protein; 372 AA.

XX AC AAB88335;

DT 23-MAY-2001 (first entry)

XX Human membrane or secretory protein clone PSEC0045.

DE Human; secretory protein; membrane protein; vaccine; gene therapy;
 vw rheumatoid arthritis; diabetes.

OS Homo sapiens.

XX EP1067182-A2.

PN 10-JAN-2001.

XX 07-JUL-2000; 2000EP-0114090.

XX 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

PI WPI; 2001-093989/11.

XX N-PSDB; AAF93762.

XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -

PS Claim 1; SEQ ID 38; 609pp + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.

XX SQ Sequence 372 AA;

Query Match 92.6%; Score 1259; DB 22; Length 372;
 Best Local Similarity 83.0%; Pred. No. 8.3e-129;
 Matches 244; Conservative 4; Mismatches 8; Indels 38; Gaps 1;

QY 1 MAVLAPLIAIYVSPRLSRWLAPYIYLLSALLSAAFLLVKRLPLCHGLPTQREDGNPCD 60
 Db 1 MAVLAPLIAIYVSPRLSRWLAPYIYLLSALLSAAFLLVKRLPLCHGLPTQREDGNPCD 60

QY 61 FDMREVEILMFLSAIVMMKNRRS----- 83
 Db 61 FDMREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120

QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIIDEELERKRVTWIIEFFANWSDCQSFAPYADL 142
 Db 121 IIVFLMTCKPPLYMGPEYIKYFNDKTIIDEELERKRVTWIIEFFANWSDCQSFAPYADL 180

QY 143 SLKYNTCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLPTLILFQGGKEAMRRPQIDKKGR 202
 Db 181 SLKYNTCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLPTLILFQGGKEAMRRPQIDKKGR 240

QY 203 AVSWTFSEENVIREFNLNELYQRAKKLSKAGDNIPPEQPVASTPTTYSNGENKK 256
 Db 241 AVSWTFSEENVIREFNLNELYQRAKKLSKAGDNIPPEQPVASTPTTYSNGENKK 294

RESULT 11

ABG18889
 ID ABG18889 standard; Protein; 344 AA.

XX AC ABG18889;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18880.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0549167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS83076.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20; SEQ ID No 49248; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving

XX Homo sapiens.
XX WO9840483-A2.
XX 17-SEP-1998.
XX 12-MAR-1998; 98WO-US04858.
XX 19-DEC-1997; 97US-0068368.
XX 14-MAR-1997; 97US-0040710.
XX 14-MAR-1997; 97US-0040762.
XX 30-MAY-1997; 97US-0048100.
XX 30-MAY-1997; 97US-0048189.
XX 30-MAY-1997; 97US-0048357.
XX 30-MAY-1997; 97US-0050934.
XX 06-JUN-1997; 97US-0048970.
XX 05-SEP-1997; 97US-0057765.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
XX Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
XX Wei YF, Young PE, Zeng Z;
XX WPI: 1998-520811/44.
XX N-PSDB; AAV34294.
XX Isolated human poly:nucleotide(s) encoding secretory peptide(s) -
XX used to develop products for the diagnosis and treatment of e.g.
XX inflammation, cancers, CNS disorders or immune system disorders
XX
XX Disclosure; Page 11; 201pp; English.
XX This sequence represents a fragment of a secreted human protein encoded
XX by the nucleic acid molecule designated Gene 9 (AAV34294). The gene can
XX be used to generate fusion proteins by linking to the gene to a human
XX immunoglobulin Fc portion (e.g. AAV34277) for increasing the stability
XX of the fused protein as compared to the human protein only.
XX The invention relates to 28 novel genes and their fragments (nucleic
XX acid sequences: AAV34286-V34325; amino acid sequences AAV75196-W75235)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological
XX conditions can be diagnosed by determining the amount of the new
XX polypeptides in a sample or by determining the presence of mutations in
XX the new polynucleotides. Specific uses are described for each of the 28
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV34286 for described uses).
XX
XX Sequence 216 AA;
Query Match 67.4%; Score 916; DB 19; Length 216;
Best Local Similarity 81.9%; Pred. No. 1.2e-91;
Matches 177; Conservative 1; Mismatches 0; Indels 38; Gaps 1;
QY 55 DGNPCDFDREVEILMFLSAIVMMKNRRS----- 83
DB 1 DGNPCDFDREVEILMFLSAIVMMKNRRSITVEQHIGNIFMESKVANTILFRLDIRMGL 60
QY 84 -----MFLMTCPEPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCQSPA 136
DB 61 LYITICIVFLMTCPEPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCQSPA 120
QY 137 PIYADLSLKYNCTGLNFGKVDVGRYTDVSTRYKVTSPKLTPLTLILFOGKEAMRRPQ 196
DB 121 PIYADLSLKYNCTGLNFGKVDVGRYTDVSTRYKVTSPKLTPLTLILFOGKEAMRRPQ 180
QY 197 IDKKGRAVSWTFSEENVITREFNLNELYQRAKKLSKA 232
DB 181 IDKKGRAVSWTFSEENVITREFNLNELYQRAKKLSKA 216
RESULT 14.

ARG01002
ID AAG01002 standard; Protein; 111 AA.
XX
AC AAG01002;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5083.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX N-PSDB; AAC01008.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 5083; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 111 AA;
Query Match 39.4%; Score 536; DB 21; Length 111;
Best Local Similarity 98.0%; Pred. No. 1.7e-50;
Matches 100; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 95 MGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCQSPAFIYADLSLKYNCTGLNFG 154
DB 10 VGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCQSPAFIYADLSLKYNCTGLNFG 69
QY 155 KYDVGRTDVSTRYKVTSPKLTPLTLILFOGKEAMRRPQ 196
DB 70 KYDVGRTDVSTRYKVTSPKLTPLTLILFOGKEAMRRPQ 111
RESULT 15
AAV12372
ID AAV12372 standard; Protein; 144 AA.
XX
AC AAV12372;
XX
DT 17-JUN-1999 (first entry)
XX
XX Human 5' EST secreted protein SEQ ID NO:403.

XX	Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW	forensic; gene therapy; chromosome mapping; signal peptide;
KW	upstream regulatory sequence; cytokine activity; cell proliferation;
KW	differentiation; haematopoiesis regulation; tissue growth regulation;
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW	thrombolytic; anti-inflammatory; tumour inhibition.
XX	
OS	Homo sapiens.
XX	
PN	WO9906548-A2.
PN	
PD	11-FEB-1999..
PD	
PF	31-JUL-1998; 98WO-TB01222.
XX	
PF	01-AUG-1997; 97US-0905135.
XX	(GEST) GENSET.
E-A	
XX	Duclet A, Dumas Milne Edwards J, Lacroix B;
PI	
XX	WPI; 1999-153778/'13.
DR	N-PSDB; AAX41205.
DR	
PT	New nucleic acids encoding human secreted proteins - obtained from
PT	cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT	kidney, lung, umbilical cord, placenta and colon tissue
XX	
XX	Claim 27; Page 725; 824pp: English.
PS	
CC	AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for '
CC	human secreted proteins', and encode the proteins given in AAY12261 to
CC	AAY12514, respectively. The proteins given represent the signal peptide
CC	and an N-terminal fragment of a secreted protein. The nucleic acid
CC	sequences can be used for producing secreted human gene products. They
CC	can also be used to develop products for diagnosis and therapy. The
CC	proteins obtained may have cytokine activity, cell
CC	proliferation/differentiation activity, haematopoiesis regulating
CC	activity, tissue growth regulating activity, reproductive hormone
CC	regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC	thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC	activity, tumour inhibition activity or other activities. The products
CC	can be used in forensic, gene therapy and chromosome mapping procedures.
CC	The sequences can also be used for obtaining corresponding promoter
CC	sequences. The nucleic acids encoding the signal peptide can be used for
CC	directing extracellular secretion of a polypeptide or the insertion of a
CC	polypeptide into a membrane, or importing a polypeptide into a cell.
XX	
XX	
SO	Sequence 144 AA;
	Query Match 37.8%; Score 514; DB 20; Length 144;
	Best Local Similarity 71.5%; Pred. No. 6.e-48;
	Matches 103; Conservative 1; Mismatches 2; Indels 38; Gaps
OY	1 MAVLAPLIAIVSYVPLSRWLQAOPYLLSALLSAFLIVRKLPPLCHGLPTQREDGNPCD 600
DB	 1 MAVLAPLIAIVSYVPLSRWLQAOPYLLSALLSAFLIVRKLPPLCHGLPTQREDGNXC 600
OY	61 FDWREVEILMFLSAIYVMKNRRS-----833
DB	61 FDWREVEILMFLSAIYVMKNRRSITVEQHIGNIFMESKVANTILFRDLRMGLLIYTLIC 120
OY	84 -MFLMTCKPPLYMGPEYIKFYFNDK 106 : 121 IVFLMTCKPPLYMGPEYIKFYFNDK 144
DB	

Search completed: July 9, 2003, 11:55:30
Job time : 75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 08:31:01 ; Search time 269 Seconds
(without alignments)
13018.057 Million cell updates/sec

Title: US-09-954-846-4
Perfect score: 1555
Sequence: 1 AGGGAGCGGGCGGAGACC.....TAATAAAGATTGGGATTA 1555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1520	97.7	1527	21 AAA15991	Human protein clon
2	1511	97.2	1598	21 AAF16316	Human prostate can
3	1429	91.9	1730	22 ABA09248	Human CGI-31 prote
4	1381.6	88.8	1699	19 AAV34294	Human secreted pro
5	1370.2	88.1	1631	22 AAF93762	Human cDNA encoding
6	1370	88.1	1620	22 AAS45940	Human DNA encoding
7	1368	88.0	1618	20 AAX52251	Protein PRO270 cDN
8	1361.8	87.6	1852	19 AAV34316	Human secreted pro
9	1361.2	87.5	1659	24 ABL49614	Prostaglandin E2 (

10	1359	87.4	1621	22 AAF72409	Human PRO270 cDNA
11	868	55.8	2297	22 ABA15531	Human nervous syst
12	868	55.8	3369	22 ABA15532	Human nervous syst
13	868	55.8	3369	22 ABA15533	Human nervous syst
14	774	49.8	774	21 AAA15981	Human protein clon
15	613	39.4	2030	23 AAS83076	DNA encoding novel
16	521	33.5	774	21 AAZ42527	Human 5' EST isola
17	463	29.8	726	22 AAF93955	Primer specific fo
18	441.2	28.4	521	16 AAT22413	Human gene signatu
19	348.4	22.4	439	21 AAA11964	Human secreted exp
20	308	19.8	469	21 AAZ43018	Human 5' EST isola
21	305.6	19.7	452	20 AAX41203	Human secreted pro
22	305.2	19.6	461	21 AAC01008	Human secreted pro
23	287.2	18.5	430	20 AAX41202	Human secreted pro
24	261	16.8	307	23 AAS83053	DNA encoding novel
25	261	16.8	444	20 AAX41205	Human secreted pro
26	258.4	16.6	307	24 ABR45061	CDNA encoding colo
27	252.2	16.2	5423	23 AAS83062	DNA encoding novel
28	252.2	16.2	5423	24 ABL68483	Kidney cancer rela
29	246	15.8	1895	23 AAS83073	DNA encoding novel
30	230.6	14.8	531	21 AAZ43016	Human 5' EST isola
31	221	14.2	756	23 ABV21932	Human prostate exp
32	221	14.2	756	23 ABV27764	Human prostate exp
33	200.4	12.9	274	21 AAA45185	Human secreted exp
34	144.6	9.3	153	22 AAL19490	Human breast cance
35	144.6	9.3	433	22 AAL09103	Human breast cance
36	141.8	9.1	370	23 ABV02989	Human prostate exp
37	110.2	7.1	1409	22 ABA15530	Human nervous syst
38	95.8	6.2	2919	23 ABL17144	Drosophila melanog
39	92.2	5.9	919	23 ABL17145	Drosophila melanog
40	89.2	5.7	367	22 ABA11541	Human nervous syst
41	84.4	5.4	375	23 ABV33303	Human prostate exp
42	84.4	5.4	375	23 ABV42226	Human prostate exp
43	84.4	5.4	389	23 ABV12158	Human prostate exp
44	50	3.2	50	20 AAX52405	Probe used to iso
45	50	3.2	50	22 AAF72563	Human PRO polypept

ALIGNMENTS

RESULT 1

AAAL5991
ID AAA15991 standard; cDNA; 1527 BP.

XX AC
XX AC
XX AC

DT 12-JUN-2000 (first entry)

XX DE Human protein clone HP10392 full length coding sequence.

XX DE Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW coagulation disorder; myocardial infarction; inflammatory condition;
KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW nephritis; therapy; ss.

XX OS Homo sapiens.

XX PN WO200005367-A2.

XX PD 03-FEB-2000.

XX PF 22-JUL-1999; 99WO-JP03929.

XX PR 24-JUL-1998; 98JP-0208820.

XX PR 07-AUG-1998; 98JP-0224105.

QY 1355 CTTTAAACCAAGTTTATCATTCAGACACATATTGACACCCCTATTGTTGGGG 1414
 DB 1321 CTTTAAACCAAGTTTATCATTCAGACACATATTGACACCCCTATTGTTGGGG 1380
 QY 1415 ATTGAGAAGGGTGAATAGAGGCTTGAGACTTTCCCTTTGTTGGTAGGACTTGGAGGAGA 1474
 DB 1381 ATTGAGAAGGGTGAATAGAGGCTTGAGACTTTCCCTTTGTTGGTAGGACTTGGAGGAGA 1440
 QY 1475 AATCCCTGGACTTTCACCTAACCTCTGACATCTCCACACCCAGTTGATGGCTTTCC 1534
 DB 1441 AATCCCTGGACTTTCACCTAACCTCTGACATCTCCACACCCAGTTGATGGCTTTCC 1500
 QY 1535 GTAATAAAAGATTGGGATT 1554
 DB 1501 GTAATAAAAGATTGGGATT 1520

RESULT 2
 AAF16316
 ID AAF16316 standard; cDNA; 1598 BP.
 XX AC
 XX AAF16316;
 DT 13-MAR-2001 (first entry)
 XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:751.
 DE Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
 XX neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
 KW vulnary; gastrointestinal; nephrotropic; antinefactive; gynaecological;
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
 KW wound; infectious disease; ss.
 XX Homo sapiens.
 OS
 XX WO200055174-A1.
 PN 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05988.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PR (ROSE/) ROSEN C A.
 ..Y
 .A
 PI Rosen CA, Ruben SW;
 XX WPI; 2000-587513/55.
 DR P-PSDB; AAB57113.
 XX Prostate cancer associated gene sequences, referred to as prostate
 PT cancer antigens, useful for treatment, prevention, and diagnosis of
 PT disorders such as prostate cancer -
 XX
 PS Claim 1; Page 1172; 2338pp; English.
 XX AAF15566 to AAF16505 encode the human prostate cancer associated
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
 CC The prostate cancer antigens can have neuroprotective, cytostatic,
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
 CC nephrotropic, antinefactive, gynaecological and antibacterial activities,
 CC and can be used in gene therapy. The prostate cancer antigen
 CC polynucleotides may be used for detection of prostate cancer, chromosome
 CC identification, as chromosome markers, and for numerous other diagnostic
 CC or research purposes. The prostate cancer antigens may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
 CC AAB57303 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 1598 BP; 437 A; 381 C; 365 G; 414 T; 1 other;
 SQ Query Match 97.2%; Score 1511; DB 21; Length 1598;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1511; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 42 CCCTTACCGCGGAAAGATGGCGTCTTGGCACCTCTAAATTTGCTCTCTGTTATTCGGTGC 101
 DB 26 CGTTCACCGCGGAAAGATGGCGTCTTGGCACCTCTAAATTTGCTCTCTGTTATTCGGTGC 85
 QY 102 CGGACTTTTACGATGGCTCGCCCACTTACCTTACCTTCTGTCGGCCCTGCTCTCTGTCG 161
 DB 86 CGGACTTTTACGATGGCTCGCCCACTTACCTTACCTTCTGTCGGCCCTGCTCTCTGTCG 145
 QY 162 CTTTCTTACTCTGAGGAAACTGCCCGCTCTGCCACGGTCTGCCACCAACCGGAAAG 221
 DB 146 CTTTCTTACTCTGAGGAAACTGCCCGCTCTGCCACGGTCTGCCACCAACCGGAAAG 205
 QY 222 ACGGTAACCCCGTGTGACTTTGACTTGGAGAAAGTGGAGATCCTGATGTTTCTCAGTGCCA 281
 DB 206 ACGGTAACCCCGTGTGACTTTGACTTGGAGAAAGTGGAGATCCTGATGTTTCTCAGTGCCA 265
 QY 282 TTGTGATGATGAAGAACCGCAGATCCATGTTCCCTGATGACGTGCAACCCCTCTATATA 341
 DB 266 TTGTGATGATGAAGAACCGCAGATCCATGTTCCCTGATGACGTGCAACCCCTCTATATA 325
 QY 342 TGGCCCTGAGTATATCAAGTACTTCAATGATAAACCATTGATGAGGAACTAGAACGGG 401
 DB 326 TGGCCCTGAGTATATCAAGTACTTCAATGATAAACCATTGATGAGGAACTAGAACGGG 385
 QY 402 ACAGAGGGTCACTTGGATTGTGGATTTCTTTGGCAATTTGGTCTAAATGACTGCCAATCAT 461
 DB 386 ACAGAGGGTCACTTGGATTGTGGATTTCTTTGGCAATTTGGTCTAAATGACTGCCAATCAT 445
 QY 462 TTGCCCTTATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGA 521
 DB 446 TTGCCCTTATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGA 505
 QY 522 AGTGGATTTGGACGGCTATAGTGTAGTAGCGGTACAAAGTGGAGCACTACACCC 581
 DB 506 AGTGGATTTGGACGGCTATAGTGTAGTAGCGGTACAAAGTGGAGCACTACACCC 565
 QY 582 TCACCAAGCAACTCCCTACCTGATCTGTTCCAAAGTGGCAAGGAGCAATGCGCGGC 641
 DB 566 TCACCAAGCAACTCCCTACCTGATCTGTTCCAAAGTGGCAAGGAGCAATGCGCGGC 625
 QY 642 CACAGATTGACAAAGAGGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCC 701
 DB 626 CACAGATTGACAAAGAGGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCC 685
 QY 702 GAGAAATTTAACTTAAATGAGCTATACCAGCGGGCCAAAGAACTATCAAGGCTGGAGACA 761
 DB 686 GAGAAATTTAACTTAAATGAGCTATACCAGCGGGCCAAAGAACTATCAAGGCTGGAGACA 745
 QY 762 ATATCCCTGAGGACAGCCTGTGGCTTCAACCCCAACAGCTGTACATGGGAAACA 821
 DB 746 ATATCCCTGAGGACAGCCTGTGGCTTCAACCCCAACAGCTGTACATGGGAAACA 805
 QY 822 AGAAGGATAAATAAGATCCTCCTTGGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 881
 DB 806 AGAAGGATAAATAAGATCCTCCTTGGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 865
 QY 882 TTCCATTAACCAAGCCTGAGGCTGACGCTTTTATTTATTTATTTTCCCTTTTGGCTGTGAC 941
 DB 866 TTCCATTAACCAAGCCTGAGGCTGACGCTTTTATTTATTTATTTTCCCTTTTGGCTGTGAC 925
 QY 942 TGGTGGGCGAGCATGACGCTTCTGATTTTAAAGAGGCACTAGGAAATTTCTAGGCAACC 1001
 DB 926 TGGTGGGCGAGCATGACGCTTCTGATTTTAAAGAGGCACTAGGAAATTTCTAGGCAACC 985
 QY 1002 CTACAGGAAGGCGCTGCCATGCTGTGGCAACTGTTTCACTGGAGCAAGAAAGAGATCTCA 1061

Qy	242	GACTGGAGAAAGTGGAGATCCTGATCTTTCTCAGTGCCCATTTGTGATGATGAAGAACC	301
Db	1447	GACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCCCATTTGTGATGATGAAGAACC	1388
Qy	302	AGATCCA-----	308
Db	1387	AGATCCATCCTGTGGAGCAACATATAGGCAACATTTTTCATGTTTGTAAAGTGGCAAC	1328
Qy	309	-----	308
Db	1327	ACAATTCTTTTTCOCGTTTGGATATTCGATGGGCTACTTTACATCACATCTCGATA	1268
Qy	309	-TGTTCTGATGACGTCAAAACCCCCCTATATATGGGCCCTCAGTATATCAAGTACTTC	367
Db	1267	GTGTTCTGATGACGTCAAAACCCCCCTATATATGGGCCCTCAGTATATCAAGTACTTC	1208
	368	AATGATAAAACCATTGATGAGGAACATAGAACGGGACAGAGGTCACCTTGGATTTGGAG	427
Db	1207	AATGATAAAACCATTGATGAGGAACATAGAACGGGACAGAGGTCACCTTGGATTTGGAG	1148
Qy	428	TTCTTTGCCAATTGGTCTAATGACTGCCAATCATTTGCCCTATCTATGCTGACCTCTCC	487
Db	1147	TTCTTTGCCAATTGGTCTAATGACTGCCAATCATTTGCCCTATCTATGCTGACCTCTCC	1088
Qy	488	CTTAAATACAACTCTACAGGCTAAATTTTGGGAAGTGGATGTTGGACCTATACTGAT	547
Db	1087	CTTAAATACAACTCTACAGGCTAAATTTTGGGAAGTGGATGTTGGACCTATACTGAT	1028
Qy	548	GTTAGTAGCGGTTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCCGTATC	607
Db	1027	GTTAGTAGCGGTTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCCGTATC	968
Qy	608	CTGTTTCCAAGTGGCAAGGAGCAATCGCGGGGCCACAGATTGACAAAGAAGACGGGCT	667
Db	967	CTGTTTCCAAGTGGCAAGGAGCAATCGCGGGGCCACAGATTGACAAAGAAGACGGGCT	908
Qy	668	GTCTCATGAGCCTCTCTGAGGGAATGTGATCGGAAATTTAACTTAAATGAGCTATAC	727
Db	907	GTCTCATGAGCCTCTCTGAGGGAATGTGATCGGAAATTTAACTTAAATGAGCTATAC	848
Qy	728	CAGCGGGCCAAAGAACTATCAAGGCTGGAGACAAATATCCCTGAGGAGCAGCCTGTGGCT	787
Db	847	CAGCGGGCCAAAGAACTATCAAGGCTGGAGACAAATATCCCTGAGGAGCAGCCTGTGGCT	788
	788	TCAACCCCCACACAGTGTGAGATGGGGAACAAAGAAGGATAAATGAATCCTCACTTT	847
Db	787	TCAACCCCCACACAGTGTGAGATGGGGAACAAAGAAGGATAAATGAATCCTCACTTT	728
Qy	848	GGCAGTCTTCTCTCTCTGTCATTCACAGCTCTTTCCATAACCAAGCCTGAGGCTGC	907
Db	727	GGCAGTCTTCTCTCTCTGTCATTCACAGCTCTTTCCATAACCAAGCCTGAGGCTGC	668
Qy	908	AGCCTTTTATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCTTCTGA	967
Db	667	AGCCTTTTATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCTTCTGA	608
Qy	968	TTTTAAAGAGGCATCTAGGAAATTTGTGAGGCACCCCTACAGAAGGCCCTGCCATGCTGTGG	1027
Db	607	TTTTAAAGAGGCATCTAGGAAATTTGTGAGGCACCCCTACAGAAGGCCCTGCCATGCTGTGG	548
Qy	1028	CCAACCTTTTCTGAGCAGGAAGAGATCTCATAGGAGGGGGGAATGTTTCCC	1087
Db	547	CCAACCTTTTCTGAGCAGGAAGAGATCTCATAGGAGGGGGGAATGTTTCCC	488
Qy	1088	TCCAAGCTTGGGTCAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGGTTTCT	1147
Db	487	TCCAAGCTTGGGTCAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGGTTTCT	428
Qy	1148	CCATGAACCTCTGTGGTTTCATCATTTCCCTTCTTGTAGTTGACCTGCACAGCTTGTGTTAGCC	1207
Db	427	CCATGAACCTCTGTGGTTTTCATCATTTCCCTTCTTGTAGTTGACCTGCACAGCTTGTGTTAGCC	368

QY	1208	TAGATTAAACCCCTAAGCTAAGATGCTGGGTATAGAACGCTAAGAAATTTTCCCCCAAGGA	1267
DB	367	TAGATTAAACCCCTAAGCTAAGATGCTGGGTATAGAACGCTAAGAAATTTTCCCCCAAGGA	308
QY	1268	CTCTTGCTTCTCTTAAGCCCTCTGGCTTCGTTTATGCTCTTCAATTAAGAGTATAGCCCTA	1327
DB	307	CTCTTGCTTCTCTTAAGCCCTCTGGCTTCGTTTATGCTCTTCAATTAAGAGTATAGCCCTA	248
QY	1328	ACTTTGTGCTAGTCTTAAGGAGAGAAACCTTTAACCCACAAAAGTTTTATCATTTGAAGACAA	1387
DB	247	ACTTTGTGCTAGTCTTAAGGAGAGAAACCTTTAACCCACAAAAGTTTTATCATTTGAAGACAA	188
QY	1388	TATTGAACAACCCCTATTTTGTGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTT	1447
DB	187	TATTGAACAACCCCTATTTTGTGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTT	128
QY	1448	CCTTTGTGCTAGGACTTTGGAGAGAGAAATCCCTTGACCTTTCACCTAACCTCTTGACATA	1507
DB	127	CCTTTGTGCTAGGACTTTGGAGAGAGAAATCCCTTGACCTTTCACCTAACCTCTTGACATA	68
QY	1508	CTCCCCACACCCAGTTGATGCTTTCGCTTAATAAAAAGATTGGGATT	1554
DB	67	CTCCCCACACCCAGTTGATGCTTTCGCTTAATAAAAAGATTGGGATT	21
RESULT 4			
AAV34294			
ID	AAV34294 standard; DNA; 1599 BP.		
XX			
AC	AAV34294;		
XX			
DT	29-JAN-1999 (first entry)		
XX			
DE	Human secreted protein gene 9 clone HJBCD89.		
XX			
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
OS	Homo sapiens.		
XX			
PN	WO9840483-A2.		
XX			
PD	17-SEP-1998.		
XX			
PF	12-MAR-1998; 98WO-US04858.		
XX			
PR	19-DEC-1997; 97US-0068368.		
PR	14-MAR-1997; 97US-0040710.		
PR	14-MAR-1997; 97US-0040762.		
PR	30-MAY-1997; 97US-0048100.		
PR	30-MAY-1997; 97US-0048189.		
PR	30-MAY-1997; 97US-0048357.		
PR	30-MAY-1997; 97US-0050934.		
PR	06-JUN-1997; 97US-0048970.		
PR	05-SEP-1997; 97US-0057765.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;		
PI	Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;		
PI	Wei YF, Young PE, Zeng Z;		
XX			
DR	WPI: 1998-520811/44.		
DR	P-PSDB; AAW75204.		
XX			
PT	Isolated human poly:nucleotide(s) encoding secretory peptide(s) -		
PT	used to develop products for the diagnosis and treatment of e.g.		

Isolated human polynucleotide(s) encoding secretory peptide(s) used to develop products for the diagnosis and treatment of e.g.

inflammation, cancers, CNS disorders or immune system disorders

Claim 1; page 121-122; 201pp; English.

This sequence represents a nucleic acid molecule which encodes a secreted human protein. The gene number, and the clone it is derived from, are detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AAV34277) for increasing the stability of the fused protein as compared to the human protein only.

The invention relates to 28 novel genes and their fragments (nucleic acid sequences: AAV34286-V34325; amino acid sequences AAW5196-WF5235) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 28 polynucleotides, based on which tissues they are most highly expressed in (see AAV34286 for described uses).

Sequence 1699 BP; 465 A; 400 C; 383 G; 449 T; 2 other;

Try Match 88.8%; Score 1381.6; DB 19; Length 1699;

t Local Similarity 92.3%; Pred. No. 0;

ches 1511; Conservative 1; Mismatches 11; Indels 114; Gaps 1;

32 CGAGCAGTGGCCGTTACGGCCGAAAGATGGCGGTC TTGGCACCTCTAATTGCTCTCGTG 91

92 TATTCGGTGCCGGCGACTTTCACGATGGCTCGCCCAACCTTACTACCTTCTGTGGCCCTG 151

63 TATTCGGTGCCGCGACTTTCACGATGGCTCGCCCAACCTTACTACCTTCTGTCGGCCCTG 122

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

123 CTCCTGCTGGCTTCCCTACTCGTGAGGAACTGCCGCCGCTCTGCCACGGTCTGCCACC 182

212 CAACGCGAAGACGGTAACCCGTGTGACTTTTGACTGGAGAGAAGTGGAGATCCTGATGTT 271

183 CAACGCCAAGACCGGTAAACCCGTTGTGACTTTGACTGGAGAGAAGTGGAGATCCTGATGTTT 242

272 CTCAGTCCATTGTGATGATGAGACCGCAGATCCA----- 308

[illegible]

243 C C A G I G C C A T T G T G A T G A I G A A G A A C C G C A G A T C C A T C A C T G T G G A G C A A C A T A T A G G C 302

309 ----- 308

303 AACATTTTCATGTTTAGTAAAGTGGCCAACACAAATTCCTTTCTTCCGCTTGGATATTCGC 362

309 ----- 327

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

363 ATGGGCC TACTT TACAT CACACT CTGCA TAGTG TCC TGATGAC GTGCA AACCCCC CCA 422

338 TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCATTTGATGAGGAAGTAGAA 397

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

398 C G G G A C A A G A G G G T C A C T T G G A T T G T G G A G T T C T T T G C C A A T T G G T C T A A T G A C T G C C A A 457

483 CGGGACAAGAGGGTCACTTGGATTGTGGAGTTCTTTGCCAATTGGTCTAATGACTGCCAA 542

A58 TCAATTGCCCCCTATTCTATTGCTGACCTTCCTCTTATAATACATCTCTACCCCTATAATAT

RESULT 5

AAAF93762

ID AAF93762 standard; cDNA; 1631 BP.

XX 55

AC
YY
AAE93/62;

23-MAY-2001 (first entry)

Qy	224	GGTAAACCCGCTGACATTTTGACTCGAGAGAAGTGGAGATCCTGATGTTTCTCACTGCGCATTT	283
Db	181	GGTAAACCCGCTGACATTTTGACTCGAGAGAAGTGGAGATCCTGATGTTTCTCACTGCGCATTT	240
Qy	284	GTGATGATGAAGAACCGCAGATCCA-----	308
Db	241	GTGATGATGAAGAACCGCAGATCCA-----	300
Qy	309	-----	308
Db	301	TTTGTAGTAAAGTGCCCAACAAATCTTTTCTTCGCTTGATATTCGCATGGGCCTACTT	360
Qy	309	-----TGTTCTCTGATCAGCTGCAAAACCCCTCTATATATGGGCCCC	349
Db	361	TACATCACACTCTGCATATGATGTTTCTCTGATGACGTGCAAAACCCCTCTATATATGGGCCCC	420
Qy	350	GAGTATATCAAGTACTTCAATGATAAACCATGATGAGGAACCTAGAACGGGACAAGAGG	409
Db	421	GAGTATATCAAGTACTTCAATGATAAACCATGATGAGGAACCTAGAACGGGACAAGAGG	480
Qy	410	GTCACTTGGATTGTGGAGTTCTTTTGGCAATTTGTTCTAAATGACTGCCCAATCATTTTGGCCCCCT	469
Db	481	GTCACTTGGATTGTGGAGTTCTTTTGGCAATTTGTTCTAAATGACTGCCCAATCATTTTGGCCCCCT	540
Qy	470	ATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGAT	529
Db	541	ATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGAT	600
Qy	530	GTTTGGAGCGCTACTGATGTTAGTACGCGGTACAAAGTGAGCACATCACCCTTCACCAAG	589
Db	601	GTTTGGAGCGCTACTGATGTTAGTACGCGGTACAAAGTGAGCACATCACCCTTCACCAAG	660
Qy	590	CAACTCCCTACCCTGATCTCTTCCAAAGTGGCAAGGAGGCAATGCGCGCGGCCACAGATT	649
Db	661	CAACTCCCTACCCTGATCTCTTCCAAAGTGGCAAGGAGGCAATGCGCGCGGCCACAGATT	720
Qy	650	GACAAGAAAGGAGGGGTGTCTCATGAGACCTTCTCTGAGGAGATGTGATCCGAGAAATTT	709
Db	721	GACAAGAAAGGAGGGGTGTCTCATGAGACCTTCTCTGAGGAGATGTGATCCGAGAAATTT	780
Qy	710	AACTTAAATGAGCTATACCAAGCGGGCCAAAGAACTATCAAAGGCTGGAGACAATATCCCT	769
Db	781	AACTTAAATGAGCTATACCAAGCGGGCCAAAGAACTATCAAAGGCTGGAGACAATATCCCT	840
Qy	770	GAGGAGCAGCCTGTGGCTTCAACCCCCACACAGTGTCTGATGGGGAACAAAGAAAGGAT	829
Db	841	GAGGAGCAGCCTGTGGC-TCAACCCCCACACAGTGTCTGATGGGGAACAAAGAAAGGAT	899
Qy	830	AAATAAGATCCTCACTTTGGCAGTGTCTCTCTCTGATTAATCCAGGCTCTTTCCATATA	889
Db	900	AAATAAGATCCTCACTTTGGCAGTGTCTCTCTCTGATTAATCCAGGCTCTTTCCATATA	959
Qy	890	CCACAGCCCTGAGGCTGACGCTTTTATGTTTCCCTTTGGCTGCTGACTGGGTGGG	949
Db	960	CCACAGCCCTGAGGCTGACGCTTTTATGTTTCCCTTTGGCTGCTGACTGGGTGGG	1019
Qy	950	GCAGCTGACGCTTCTGATTTTAAAGAGGCACTAGGGAATGTCAGGCACCCCTACAGGA	1009
Db	1020	GCAGCTGACGCTTCTGATTTTAAAGAGGCACTAGGGAATGTCAGGCACCCCTACAGGA	1079
Qy	1010	AGGCCTGCCATGCTGTGGCCAACTGTTTCTACTGGAGCAAGAAAGAGATCTCATAGACAGG	1069
Db	1080	AGGCCTGCCATGCTGTGGCCAACTGTTTCTACTGGAGCAAGAAAGAGATCTCATAGACAGG	1139
Qy	1070	AGGGGGAATGGTTTCCCTCCAAGCTTGGGCTGAGTGTAACTGCTTATCAGCTATTCA	1129
Db	1140	AGGGGGAATGGTTTCCCTCCAAGCTTGGGCTGAGTGTAACTGCTTATCAGCTATTCA	1199
Qy	1130	GACATCCCATGGTTTCTCCATGAAACTCTGTGGTTTTCATCATTCCTTCTAGTTGACCT	1189
Db	1200	GACATCCCATGGTTTCTCCATGAAACTCTGTGGTTTTCATCATTCCTTCTAGTTGACCT	1259
Qy	1190	GCACAGCTTGGTTAGACCTAGATTTAAACCCCTAAAGGTAAGTGTGGGGTATAGAACGCTA	1249

Db 1260 GCACAGCTTGGTTAGACCTAGATTAAACCTAAGGTAAGATGCTGGGGTATAGAACGCTA 1319
QY 1250 AGAATTTTCCCAAGACACTTCTGCTTCCCTTAAGCCCTTCTGGCTTCTGTTTATGCTTTC 1309
Db 1320 AGAATTTTCCCAAGACACTTCTGCTTCCCTTAAGCCCTTCTGGCTTCTGTTTATGCTTTC 1379
QY 1310 ATTAAGAGTATAAGCCCTAAGCTTTGCTGCTAGTCTTAAGGAGAAACCTTTTAACCAACAAGT 1369
Db 1380 ATTAAGAGTATAAGCCCTAAGCTTTGCTGCTAGTCTTAAGGAGAAACCTTTTAACCAACAAGT 1439
QY 1370 TTTTATCATTTGAAGACAATATTGAACAACCCCTTATTTGTGGGGATTGAGAAGGGTGA 1429
Db 1440 TTTTATCATTTGAAGACAATATTGAACAACCCCTTATTTGTGGGGATTGAGAAGGGTGA 1499
QY 1430 ATAGAGGCTTGAGACTTCTCTTTGTGTGGTAGGACTTGGAGGAGAAATCCCTGGACTTT 1489
Db 1500 ATAGAGGCTTGAGACTTCTCTTTGTGTGGTAGGACTTGGAGGAGAAATCCCTGGACTTT 1559
QY 1490 CACTAACCCTCTGACATACCTCCACACCCAGTTGATGGCTTTCCGTAATAAAGAGATTG 1549
Db 1560 CACTAACTCTGACATACCTCCACACCCAGTTGATGGCTTTCCGTAATAAAGAGATTG 1619
1550 GGATT 1554
1620 GGATT 1624

RESULT 6
AAS45940
ID AAS45940 standard; cDNA; 1620 BP.
AC AAS45940;
XX
DT 18-DEC-2001 (first entry)
XX Human DNA encoding PRO polypeptide sequence #16.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.

01-MAR-2000; 2000WO-US05601.
02-MAR-2000; 2000WO-US05841.
03-MAR-2000; 2000US-187202P.
06-MAR-2000; 2000US-186968P.
14-MAR-2000; 2000US-189320P.
14-MAR-2000; 2000US-189328P.
15-MAR-2000; 2000WO-US06884.
21-MAR-2000; 2000US-190828P.
21-MAR-2000; 2000US-191007P.
21-MAR-2000; 2000US-191048P.
21-MAR-2000; 2000US-191314P.
28-MAR-2000; 2000US-192655P.
29-MAR-2000; 2000US-193032P.
29-MAR-2000; 2000US-193053P.
30-MAR-2000; 2000WO-US08439.
04-APR-2000; 2000US-19449P.
04-APR-2000; 2000US-194647P.
11-APR-2000; 2000US-195975P.
11-APR-2000; 2000US-196000P.
11-APR-2000; 2000US-196187P.
11-APR-2000; 2000US-196690P.

11-APR-2000; 2000US-196820P.
18-APR-2000; 2000US-198121P.
18-APR-2000; 2000US-198585P.
25-APR-2000; 2000US-199397P.
25-APR-2000; 2000US-199550P.
25-APR-2000; 2000US-199654P.
03-MAY-2000; 2000US-201516P.
17-MAY-2000; 2000WO-US13705.
22-MAY-2000; 2000WO-US14042.
30-MAY-2000; 2000WO-US14941.
02-JUN-2000; 2000WO-US15264.
05-JUN-2000; 2000US-209832P.
28-JUL-2000; 2000WO-US20710.
22-AUG-2000; 2000US-0644848.
24-AUG-2000; 2000WO-US23328.
08-NOV-2000; 2000WO-US30952.
01-DEC-2000; 2000WO-US32678.
20-DEC-2000; 2000WO-US34956.
(GETH) GENENTECH INC.
Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
WPI: 2001-602745/68.
P-PSDB: RAU29039.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the presence of tumours, such as prostate and breast tumours, in mammals and to screen for modulators of the compounds -
Claim 2; Fig 31; 774pp; English.
Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR primers for PRO polypeptides of the invention. The sequences of the invention can be used to detect the presence of a tumour in a mammal by comparing the level of expression of a PRO polypeptide in a test sample of cells from the animal and a control sample of normal cells, whereby a higher level of expression in the test sample indicates the presence of a tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats and rabbits but are preferably human. The polypeptides can be used to stimulate tumour necrosis factor (TNF) alpha release from human blood, when contacted with it. A specific polypeptide can be used to stimulate the proliferation or differentiation of chondrocyte cells. The PRO proteins can be used to determine the presence of tumours and also susceptibility to tumour development, particularly adrenal, lung, colon, breast, prostate, rectal, cervical, or liver tumours, in mammalian subjects. The oligonucleotide probes specific for the PRO nucleic acids can be used for genetic analysis of individuals with genetic disorders.

Sequence 1620 BP; 407 A; 394 C; 373 G; 442 T; 4 other;
Query Match 88.1%; Score 1370; DB 22; Length 1620;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1494; Conservative 0; Mismatches 4; Indels 114; Gaps 1;
QY 57 AGATGGGGTCTTGGACCTCTAATGCTCTCGTGTATTCGGTCCCGGACTTTCACGAT 116
Db 1 AGATGGGGTCTTGGACCTCTAATGCTCTCGTGTATTCGGTCCCGGACTTTCACGAT 60
QY 117 GGCTCGCCCAACCTTACTACTCTGTCGGCCCTGCTCTGCTCCCTTCTACTCTGTA 176
Db 61 GGCTCGCCCAACCTTACTACTCTGTCGGCCCTGCTCTGCTCCCTTCTACTCTGTA 120
QY 177 GGAACCTGCCCGCGCTCTGCGACGGTCTGCCACCAACGAGACGTAACCGTGTG 236
Db 121 GGAACCTGCCCGCGCTCTGCGACGGTCTGCCACCAACGAGACGTAACCGTGTG 180
QY 237 ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCCATTTGATGATGAAGA 296
Db 181 ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCCATTTGATGATGAAGA 240
QY 297 ACCGCATGCA----- 308

Db	241	ACCGCAGATCCATCACTGTGGAGCAACATATAGCAACATATTTTCATGTTTAGTAAAGTGG	300
Qy	309	-----	308
Db	301	CCAACACAAATCTTTTCCTCCGTTGGATATTCGCAATGGCCCTACTTTACATCACACTCT	360
Qy	309	-----TGTTCCCTGATGACGTGCAAAACCCCCCTATATATATGGGCCCTGAGTATATCAAGT	362
Db	361	GCATAGTGTTCCTGATGACGTGCAAAACCCCTATATATATGGGCCCTGAGTATATCAAGT	420
Qy	363	ACTTCAATGATAAAACCAATGATGAGAACTAGAACGGGCAAGAGGGTCACTTGGATTG	422
Db	421	ACTTCAATGATAAAACCAATGATGAGAACTAGAACGGGCAAGAGGGTCACTTGGATTG	480
Qy	423	TGAGTCTTTGGCAATTTGCTTAATGACATGCGCAATCATTTGCCCTATCTATGCTGACC	482
Db	481	TGGAGTCTTTGGCAATTTGCTTAATGACATGCGCAATCATTTGCCCTATCTATGCTGACC	540
Qy	483	TCCTCCCTTAATAACAACGTACAGGGCTAAATTTTGGGAAGGTGGATTTGGAGCGCTATA	542
Db	541	TCCTCCCTTAATAACAACGTACAGGGCTAAATTTTGGGAAGGTGGATTTGGAGCGCTATA	600
Qy	543	CTGATGTTAGTAGCGGTACAAAAGTAGACACATCACCCCTACCAAGCAACTCCCTACCC	602
Db	601	CTGATGTTAGTAGCGGTACAAAAGTAGACACATCACCCCTACCAAGCAACTCCCTACCC	660
Qy	603	TGATCTGTTCCAAAGTGGCAAGAGGCAATCGCGCGCCACACATTCACAGAAGAGAC	662
Db	661	TGATCTGTTCCAAAGTGGCAAGAGGCAATCGCGCGCCACACATTCACAGAAGAGAC	720
Qy	663	GGGCTGCTCATGACCTTCTCTGAGGAGATGTGATCCGAGAAATTTAACTTTAAATGAGC	722
Db	721	GGGCTGCTCATGACCTTCTCTGAGGAGATGTGATCCGAGAAATTTAACTTTAAATGAGC	780
Qy	723	TATACCAAGCGGGCCAAAGAACTATCAAAAGCTGGAGACAATATCCCTGAGGAGAGCGCTG	782
Db	781	TATACCAAGCGGGCCAAAGAACTATCAAAAGCTGGAGACAATATCCCTGAGGAGAGCGCTG	840
Qy	783	TGGCTTCAACCCCAACACAGTGTACAGTGGGGGAAACAGAGGATAAATAAGATCCTC	842
Db	841	TGGCTTCAACCCCAACACAGTGTACAGTGGGGGAAACAGAGGATAAATAAGATCCTC	900
Qy	843	ACTTTGACGTGCTCTCTCTCTCAATTCACAGCTCTTCCATACCAACAGCCTGAG	902
Db	901	ACTTTGACGTGCTCTCTCTCTCAATTCACAGCTCTTCCATACCAACAGCCTGAG	960
Qy	903	GCTGACGCTTTTATTTATGTTTTCCTTTGGCTGTGACTGGGTGGGGCAGCATGCAAGT	962
Db	961	GCTGACGCTTTTATTTATGTTTTCCTTTGGCTGTGACTGGGTGGGGCAGCATGCAAGT	1020
Qy	963	TCGTATTTTAAAGAGGATCTAGGGAATTTGACAGGACCCCTACAGGAAGGCTGCCATGC	1022
Db	1021	TCGTATTTTAAAGAGGATCTAGGGAATTTGACAGGACCCCTACAGGAAGGCTGCCATGC	1080
Qy	1023	TGTGGCAACTGTTTCTACTGGAGCAAGAGAGATCTCATAGGACGGGGGAAATGGT	1082
Db	1081	TGTGGCAACTGTTTCTACTGGAGCAAGAGAGATCTCATAGGACGGGGGAAATGGT	1140
Qy	1083	TTCCCTCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG	1142
Db	1141	TTCCCTCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG	1200
Qy	1143	TTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTAGTTGACCTGCACAGCTTGGTT	1202
Db	1201	TTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTAGTTGACCTGCACAGCTTGGTT	1260
Qy	1203	AGACCTAGATTTAACCTTAAGTAAAGTGTGGGGTATAGAACGCTAAGAAATTTTCCCCC	1262
Db	1261	AGACCTAGATTTAACCTTAAGTAAAGTGTGGGGTATAGAACGCTAAGAAATTTTCCCCC	1320
Qy	1263	AAGGACCTCTGCTTCTCTTAAAGCCCTCTCTGGCTTGGTTTATGCTTCTTATTAAGATATAA	1322

Db 661 CAAGGTGGCAAGGAGCAATCGCGGCCACAGATTGACAAGAAAGGACGGCTGCTCTCA 720
Qy 674 TGGACCTTCTCTGAGGAGAATGTGATCCGAGAATTAATTAATGAGCTATACACGGG 733
Db 721 TGGACCTTCTCTGAGGAGAATGTGATCCGAGAATTAATTAATGAGCTATACACGGG 780
Qy 734 GCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACC 793
Db 781 GCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACC 840
Qy 794 CCCACACAGTGTGAGTGGGGAACAAAGAGATAAATAGATCCCTCACTTTGGCAGT 853
Db 841 CCCACACAGTGTGAGTGGGGAACAAAGAGATAAATAGATCCCTCACTTTGGCAGT 900
Qy 854 GCTTCTCTCTGCTCAATTCAGGCTCTTCCATTAACACACAGCCTGAGGCTGACGCTT 913
Db 901 GCTTCTCTCTGCTCAATTCAGGCTCTTCCATTAACACACAGCCTGAGGCTGACGCTT 960
Qy 914 TTATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCGCTTCTGATTTAA 973
Db 961 TTATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCGCTTCTGATTTAA 1020
Qy 974 AGAGGCATCTAGGGAATGTGAGGACACCTTACAGGAAGGCTGCCATGCTGTGGCAACT 1033
Db 1021 AGAGGCATCTAGGGAATGTGAGGACACCTTACAGGAAGGCTGCCATGCTGTGGCAACT 1080
Qy 1034 GTTTCACCTGGAGCAAGAAAGATCTCATAGACGGAGGGGAAATGTTTCCCTCCCAAG 1093
Db 1081 GTTTCACCTGGAGCAAGAAAGATCTCATAGACGGAGGGGAAATGTTTCCCTCCCAAG 1140
Qy 1094 CTTGGGTCAGTGTCTTAAGTCTATCAGCTATTCAGACATCTCCATGTTCTCCATGA 1153
Db 1141 CTTGGGTCAGTGTCTTAAGTCTATCAGCTATTCAGACATCTCCATGTTCTCCATGA 1200
Qy 1154 AACTCTGTGTTTCATCATCTCTTCTAGTTGACCTGCACAGCTTGGTTAGACCTAGATT 1213
Db 1201 AACTCTGTGTTTCATCATCTCTTCTAGTTGACCTGCACAGCTTGGTTAGACCTAGATT 1260
Qy 1214 TAACCCCTAAGTAAAGTCTGGGTATAGACGCTAAGAAATTTCCCCCAAGGACTCTTG 1273
Db 1261 TAACCCCTAAGTAAAGTCTGGGTATAGACGCTAAGAAATTTCCCCCAAGGACTCTTG 1320
Qy 1274 CTTCCCTTAAGCCCTTCTGGCTTCTGTTTATGTTCTTCAATTAAGATTAAGCCTAAGTTG 1333
Db 1321 CTTCCCTTAAGCCCTTCTGGCTTCTGTTTATGTTCTTCAATTAAGATTAAGCCTAAGTTG 1380
Qy 1334 TCGTAGTCTTAAGGAGAAACCTTTAACCACAAAGTTTTTATCATTTGAAGACAATATTGA 1393
Db 1381 TCGTAGTCTTAAGGAGAAACCTTTAACCACAAAGTTTTTATCATTTGAAGACAATATTGA 1440
Qy 1394 ACAACCCCTATTTTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAGACTTCCCTTTG 1453
Db 1441 ACAACCCCTATTTTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAGACTTCCCTTTG 1500
Qy 1454 TGTGTAGGACTTGGAGAGAAATCCCTGACCTTCACTAACCTCTGACATATCCCC 1513
Db 1501 TGTGTAGGACTTGGAGAGAAATCCCTGACCTTCACTAACCTCTGACATATCCCC 1560
Qy 1514 ACACCCAGTGTGATGGCTTTCGGTAAATAAAGATTGGGATT 1554
Db 1561 ACACCCAGTGTGATGGCTTTCGGTAAATAAAGATTGGGATT 1601

RESULT 9

ABLA9614

ID ABL49614 standard; DNA; 1659 BP.

XX

AC ABL49614;

XX

DT 01-JUN-2002 (first entry)

XX

DE Prostaglandin E2 (PGE2) encoding DNA.

XX

Prostaglandin E2; PGE2; stomach cancer; gene; ds.

Unidentified.

Key Location/Qualifiers

CDS 31..921

/*tag= a

/product= "prostaglandin E2"

KR2001081233-A.

XX

XX

PD 29-AUG-2001.

PF 11-FEB-2000; 2000KR-0006416.

XX

PR 11-FEB-2000; 2000KR-0006416.

XX

PA (NAHE-) NAT INST HEALTH IN KOREA.

PI Kim GC, Nam MJ, Park MS;

XX

DR WPI; 2002-136917/18.

XX

PT P-PSDB; ABB06300.

XX

Novel gene expressed by prostaglandin E2 (PGE2) in stomach cancer cell

line snu-1 and its amino acid sequence -

XX

Disclosure; Page 12; 12pp; Korean.

PS

The present sequence encodes prostaglandin E2 (PGE2) which is expressed

in the stomach cancer cell line SNU-1. The PGE2 gene can be used in the

treatment of stomach cancer.

XX

Sequence 1659 BP; 426 A; 404 C; 380 G; 448 T; 1 other;

SQ

Query Match 87.5%; Score 1361.2; DB 24; Length 1659;

Best Local Similarity 92.7%; Pred No. 0;

Matches 1498; Conservative 0; Mismatches 3; Indels 115; Gaps 2;

Qy 54 AAAAGATGGCGTCTTGGCACCTCTAAATGCTCTCGTGTATTGCGTGGCGGACTTTCAC 113

Db 26 ACAAGATGGCGTCTTGGCACCTCTAAATGCTCTCGTGTATTGCGTGGCGGACTTTCAC 85

Qy 114 GATGCTCGCCCAACCTTACTACTCTCTGTCGGCCCTGCTCTGCTGCTTCTTCTACTCG 173

Db 86 GATGCTCGCCCAACCTTACTACTCTCTGTCGGCCCTGCTCTGCTGCTTCTTCTACTCG 145

Qy 174 TGAGAAACTGCGCGCGTCTGCCACGGTCTGCCACCAACCGCAAGACGGTAACCCGT 233

Db 146 TGAGAAACTGCGCGCGTCTGCCACGGTCTGCCACCAACCGCAAGACGGTAACCCGT 205

Qy 234 GTGACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCATTTGTGATGATGA 293

Db 206 GTGACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCATTTGTGATGATGA 265

Qy 294 AGAACCGCAGATCCA----- 308

Db 266 AGAACCGCAGATCCA----- 325

Qy 309 ----- 308

Db 326 TGGCCAACACAAATCTTTCTTCCGCTTGGATATTGCGATGGCGCTACTTTTACATCACAC 385

Qy 309 -----TGTTCTGATGACGTGCAAAACCCCTTATATATGGCCCTGAGTATATCA 359

Db 386 TCTGATGATGTTCTGATGAGTGCACAAACCCCTTATATATGGCCCTGAGTATATCA 445

Qy 360 AGTACTTCAATGATAAAACCACTTATGAGGAAGTAAAGCGGACAGAGGGTCACTTGA 419

Db 446 AGTACTTCAATGATAAAACCACTTATGAGGAAGTAAAGCGGACAGAGGGTCACTTGA 505

Qy 420 TTGTGGAGTCTTTGCCAATGGTCTAATGACTGCTCAATCTATTTGCCCTATCTATGCTG 479

506 TTGTGGAGTCTTTGCCAATGGTCTAATGACTGCCAATCAATTTGCCCCCTATCTATGCTG 565
480 ACCTCTCCCTTAAATACAACTGTACAGGCTAAATTTGGGAAGTGGATGTTGGACGCT 539
566 ACCTCTCCCTTAAATACAACTGTACAGGCTAAATTTGGGAAGTGGATGTTGGACGCT 625
540 ATACTGATGTTAGTACGGGTACAAGTACAGACATCACCCCTCACCAAGCACTCCCTA 599
626 ATACTGATGTTAGTACGGGTACAAGTACAGACATCACCCCTCACCAAGCACTCCCTA 685
600 CCCTGATCTGTTCCTCAAGTGCACAGGAGCAATCGCGGGCCACAGATTGACAAAGAA 659
686 CCCTGATCTGTTCCTCAAGTGCACAGGAGCAATCGCGGGCCACAGATTGACAAAGAA 745
660 GACGGGCTGTCATGAGGCTTCTCTGAGGAGAAATGTATCGAGAAATTAACCTTAATG 719
746 GACGGGCTGTCATGAGGCTTCTCTGAGGAGAAATGTATCGAGAAATTAACCTTAATG 805
720 AGCTATACCGGGGCCAAGAACTATCAAGGCTGGAGACATATCCCTGGAGGACG 779
806 AGCTATACCGGGGCCAAGAACTATCAAGGCTGGAGACATATCCCTGGAGGACG 865
780 CTGTGGCTTCAACCCGCCACAGTGTACAGTGGGAAACAAAGAGGATTAATAAGAT 839
866 CTGTGGCTTCAACCCGCCACAGTGTACAGTGGGAAACAAAGAGGATTAATAAGAT 925
840 CTACTTTGGCAGTCTCTCTCTGTCATTTCCAGGCTCTTCCATACCAACCAAGCCT 899
926 CTACTTTGGCAGTCTCTCTCTGTCATTTCCAGGCTCTTCCATACCAACCAAGCCT 985
900 GAGGCTGAGCCCTTTTATTTATGTTT - TCCCTTTGGCTGTGACTGGGTGGGCGAGATGC 958
986 GAGGCTGAGCCCTTTTATTTATGTTTCTCCCTTTGGCTGTGACTGGGTGGGCGAGATGC 1045
959 AGCTTCTGATTTTAAAGGCACTAGGCAATTTGTCAGCACCTTACAGGAGGCTGCC 1018
1046 AGCTTCTGATTTTAAAGGCACTAGGCAATTTGTCAGGCACTTACAGGAGGCTGCC 1105
1019 ATGCTGTGGCCAACTGTTTCACTGGAGCAAGAAAGATCTCATAGGAGGAGGGGAAA 1078
1106 ATGCTGTGGCCAACTGTTTCACTGGAGCAAGAAAGATCTCATAGGAGGAGGGGAAA 1165
1079 TGGTTTCCCTCAAGCTTGGGTCAGTGTGTTTAACTGCTTATCAGCTATTACAGATCTCC 1138
1166 TGGTTTCCCTCAAGCTTGGGTCAGTGTGTTTAACTGCTTATCAGCTATTACAGATCTCC 1225
1139 ATGTTTCTCCATGAACCTCTGTTTTCATCATCTCTTCTAGTTGACCTGCACAGCTT 1198
1226 ATGTTTCTCCATGAACCTCTGTTTTCATCATCTCTTCTAGTTGACCTGCACAGCTT 1285
1199 GGTAGACCTAGATTTAACCCCTAAGTAAAGTCTGGGGTATAGACGCTAAGAAATTTTC 1258
1286 GGTAGACCTAGATTTAACCCCTAAGTAAAGTCTGGGGTATAGACGCTAAGAAATTTTC 1345
1259 CCCAAGGACTCTGCTTCCCTTAAAGCCCTTGGCTTCTGTTATGTTCTTCAATAAAGT 1318
1346 CCCAAGGACTCTGCTTCCCTTAAAGCCCTTGGCTTCTGTTATGTTCTTCAATAAAGT 1405
1319 ATAAGCCTAACTTGTGCTAGTCTTAAAGGAGAACTTTTAAACCAAGAAATTTTATCAT 1378
1406 ATAAGCCTAACTTGTGCTAGTCTTAAAGGAGAACTTTTAAACCAAGAAATTTTATCAT 1465
1379 TGAAGACATATTTGAACAAACCCCTTATTTTGGGGATTGAGAGGGGTGAATAGAGGCT 1438
1466 TGAAGACATATTTGAACAAACCCCTTATTTTGGGGATTGAGAGGGGTGAATAGAGGCT 1525
1439 TGAGACTTTCTTGTGTGGTAGGACTTGGAGGAGAAATCCCTCGGACTTCTACTAACCC 1498
1526 TGAGACTTTCTTGTGTGGTAGGACTTGGAGGAGAAATCCCTCGGACTTCTACTAACCC 1585
1499 TGTGACATACTCCCAACCCAGTGTGATGCTTCCGTTAATAAAGATTTGGATT 1554
1586 TGTGACATACTCCCAACCCAGTGTGATGCTTCCGTTAATAAAGATTTGGATT 1641

RESULT 10

AAF72409

ID : AAF72409 standard; cDNA; 1621 BP.

XX

AC AAF72409;

XX

DT 24-APR-2001 (first entry)

XX

DE Human PRO270 cDNA.

XX

Human; PRO: dermatological; antipsoriatic; cytostatic; antiinflammatory;
antiParkinsonian nootropic; neuroprotective; vulnery; cardiant;
KW antiangiogenic; vasotropic; antiasthmatic; antineumatic; cancer;
KW antithratic; antinfertility; antidiabetic; antiviral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW ischaemia; inflammation; ss.

XX

OS Homo sapiens.

XX

PN WO200104311-A1.

XX

PD 18-JAN-2001.

XX

PF 22-FEB-2000; 2000WO-US04414.

XX

PR 07-JUL-1999; 99US-0143048.

XX

PR 26-JUL-1999; 99US-0145698.

XX

PR 28-JUL-1999; 99US-0146222.

XX

PR 08-SEP-1999; 99WO-US20594.

XX

PR 13-SEP-1999; 99WO-US20944.

XX

PR 15-SEP-1999; 99WO-US21090.

XX

PR 15-SEP-1999; 99WO-US21547.

XX

PR 05-OCT-1999; 99WO-US23089.

XX

PR 29-NOV-1999; 99WO-US28214.

XX

PR 30-NOV-1999; 99WO-US28313.

XX

PR 16-DEC-1999; 99WO-US30095.

XX

PR 20-DEC-1999; 99WO-US30911.

XX

PR 20-DEC-1999; 99WO-US30999.

XX

PR 05-JAN-2000; 99WO-US00219.

XX

(GETH) GENENTECH INC.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
Williams PM, Wood WI;
WPI; 2001-081051/09.
P-PSDB; AAB80248.
Sixty one nucleic acids encoding PRO polypeptides which are useful in
the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
squamous cell carcinoma) and neurodegenerative diseases (e.g.
Alzheimer's disease)
Claim 2; Fig 75; 393pp; English.
The present sequence is one of sixty one nucleic acids encoding novel
secreted and transmembrane PRO polypeptides. The PRO polypeptides are
useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
squamous cell carcinoma), gastrointestinal disorders (e.g.
enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
Parkinson's disease), wound repair, cardiovascular disorders (e.g.
endometrial bleeding angiogenesis, ischaemias such as coronary
ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
diabetes and retinal disorders such as retinitis pigmentosa.
The PRO nucleic acids have applications in molecular biology, including
use as hybridization probes, and in chromosome and gene mapping.

SQ		Sequence 1621 BP; 407 A; 394 C; 374 G; 442 T; 4 other;	
Query Match		87.4%; Score 1359; DB 22; Length 1621;	
Best Local Similarity		92.6%; Pred. No. 0;	
Matches 1494; Conservative		0; Mismatches 4; Indels 115; Gaps 2;	
Qy	57	AGATGGGGCTTGGCAGCTTAATGCTCTCGTGTATTCGGTCCGCGAGCTTTCACCAT	116
Db	1	AGATGGGGCTTGGCAGCTTAATGCTCTCGTGTATTCGGTCCGCGAGCTTTCACCAT	60
Qy	117	GGGTGCGCCCAACCTTACTACCTTCTGTGCGCCCTGCTCTCTGCTGCTTCTTCTACTCGTGA	176
Db	61	GGCTGCGCCCAACCTTACTACCTTCTGTGCGCCCTGCTCTCTGCTGCTTCTTCTACTCGTGA	120
Qy	177	GGAACCTGCGCCCTCTGCCACGGTCTGCCACCCACCGGAGAGAGCGGTAAACCGGTGTG	236
Db	121	GGAACCTGCGCCCTCTGCCACGGTCTGCCACCCACCGGAGAGAGCGGTAAACCGGTGTG	180
Qy	237	ACTTTGACTGGAGAGAGTGGAGATCCTGTATGTTCTCAGTGCCATTGTGATGATGAAGA	296
Db	181	ACTTTGACTGGAGAGAGTGGAGATCCTGTATGTTCTCAGTGCCATTGTGATGATGAAGA	240
Qy	297	ACCGCAGATCCA-----	308
Qy	241	ACCGCAGATCCATCTGTGGAGCAACATATAGCAACATTTTCATGTTTAGTAAAGTGG	300
Qy	309	-----	308
Db	301	CCAAACAAATCTTTTCTTCGGCTGGATATTCGCATGGGCTACTTTACATCACACCTCT	360
Qy	309	-----TGTTCTCATGAGCGCAAAACCCCTCTATATATGCGCTCAGTATATCAAGT	362
Db	361	GCATAGTGTCTCTGATGAGCGTGAACCCCTCTATATATGCGCTCAGTATATCAAGT	420
Qy	363	ACTTCAATGATPAAACCACTTATGAGGAATAGAACGGGACAAAGAGGTCTACTTGGATTG	422
Db	421	ACTTCAATGATPAAACCACTTATGAGGAATAGAACGGGACAAAGAGGTCTACTTGGATTG	480
Qy	423	TGGAGTCTTGGCAATTTGGTCTTAATGACTGCCAATCAATTTGCCCTATCTATGCTGACC	482
Db	481	TGGAGTCTTGGCAATTTGGTCTTAATGACTGCCAATCAATTTGCCCTATCTATGCTGACC	540
Qy	483	TCCTCCCTAAATCAACTGTACAGGGCTAAATTTTGGGAAGTGGATGTGGACGCTATA	542
Db	541	TCCTCCCTAAATCAACTGTACAGGGCTAAATTTTGGGAAGTGGATGTGGACGCTATA	600
Qy	543	CTGATGTTAGTACGCGGTACAAAGTGAACATCACCCCTCACCAAGCAACTCCCTACCC	602
Db	601	CTGATGTTAGTACGCGGTACAAAGTGAACATCACCCCTCACCAAGCAACTCCCTACCC	660
Qy	603	TGATCCTGTTCCAAAGTGGCAAGGAGCAATTCGGCGGCCACAGATTGACAAAGGAC	662
Db	661	TGATCCTGTTCCAAAGTGGCAAGGAGCAATTCGGCGGCCACAGATTGACAAAGGAGAC	720
Qy	663	GGGTGCTCTCAT- GGACCTTCTCTGAGAGATGTGATCCGAGATTTAACTTAATGAG	721
Db	721	GGGTGCTCTCATGAGGACCTCTCTGAGGAGATGTGATCCGAGATTTAACTTAATGAG	780
Qy	722	CTATACCAAGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCGCT	781
Db	781	CTATACCAAGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCGCT	840
Qy	782	GTGGCTTCAACCCCCACACAGTGTACAGATGGGAAACAAAGAGGATAAATAAGATCCT	841
Db	841	GTGGCTTCAACCCCCACACAGTGTACAGATGGGAAACAAAGAGGATAAATAAGATCCT	900
Qy	842	CACCTTGGCAGTCTCTCTCTCTCTCAATTCAGGCTCTTTCATATACCAAGCCCTGA	901
Db	901	CACCTTGGCAGTCTCTCTCTCTCTCAATTCAGGCTCTTTCATATACCAAGCCCTGA	960
Qy	902	GGCTGCGACCTTTTATTTATCTTTTCCCTTGGCTGTGACTGGGTGGGCGGACGATGAGC	961
Db	961	GGCTGCGACCTTTTATTTATCTTTTCCCTTGGCTGTGACTGGGTGGGCGGACGATGAGC	1020

RESULT 11

ABAL5531

ID ABAL5531 standard; DNA; 2297 BP.

XX ABAL5531;

DT: 23-JAN-2002 (first entry)

XX Human nervous system related polynucleotide SEQ ID NO 7862.

Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
 immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 antiparkinsonian; antiskliding; antianaemic; antiarthritic; cancer;
 antifheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 antiparasitic; immune disorder; cardiovascular disorder;
 neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX Homo sapiens.

OS WO200159063-A2.

PN 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 27-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-541565/60.
Nucleic acids encoding 3224 human nervous system antigen polypeptides,
useful for preventing, diagnosing and/or treating nervous system
cancers and metastases -
Disclosure; SEQ ID NO 7862; 1701pp + Sequence Listing; English.
The invention relates to novel genes (ABA11004-ABA21534) and proteins
(ABB14678-ABB18001) useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pt_sequences.
XX
SQ Sequence 2297 BP; 585 A; 510 C; 546 G; 656 T; 0 other;

Query Match 55.8%; Score 868; DB 22; Length 2297;
Best Local Similarity 100.0%; Pred. No. 2.8e-272;
Matches 868; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

687 AGGAGATGTGATCCGAGATTTAACTTAATGAGCTATACAGGGGCCCAAGAACTAT 746
|||||
1419 AGGAGATGTGATCCGAGATTTAACTTAATGAGCTATACAGGGGCCCAAGAACTAT 1478
|||||
747 CAAGAGCTGGAGACAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGT 806
|||||
1479 CAAGAGCTGGAGACAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGT 1538
|||||
807 CAGATGGGGAAACAAGAGGATAAATAAGATCTCTCACTTTGGCAGTGTCTCTCTCTG 866
|||||
1539 CAGATGGGGAAACAAGAGGATAAATAAGATCTCTCACTTTGGCAGTGTCTCTCTCTG 1598
|||||
867 TCAATTCAGGCTCTTCCCAATACCAAGAGCTGAGGCTGAGCTTTTATTTATGTTTT 926
|||||
1599 TCAATTCAGGCTCTTCCCAATACCAAGAGCTGAGGCTGAGCTTTTATTTATGTTTT 1658
|||||
927 CCCTTTGGCTGTACTGGGTGGGAGCAGATCGACCTTCTGATTTTAAAGAGGCACTAGG 986
|||||
1659 CCCTTTGGCTGTACTGGGTGGGAGCAGATCGACCTTCTGATTTTAAAGAGGCACTAGG 1718
|||||
987 GAATTTGCAGCAGCCCTACAGGAGGCTGCCATGCTGTGGCCCACTTTTCACTGGAGC 1046
|||||
1719 GAATTTGCAGCAGCCCTACAGGAGGCTGCCATGCTGTGGCCCACTTTTCACTGGAGC 1778
|||||
1047 AAGAAAGAGATCTCATAGGAGGAGGGGAATGGTTTCCCTCCAAGCTTGGGTCAGTGT 1106
|||||
1779 AAGAAAGAGATCTCATAGGAGGAGGGGAATGGTTTCCCTCCAAGCTTGGGTCAGTGT 1838
|||||
1107 GTTAACCTGCTTATCAGCTATTCAGACATCTCCATGGTTTCCCATGAACCTCTGTGGTTT 1166
|||||
1839 GTTAACCTGCTTATCAGCTATTCAGACATCTCCATGGTTTCCCATGAACCTCTGTGGTTT 1898
|||||
1167 CATCATTCCTTCTAGTGTAGCTCCAGCTGCTGTAGACCTAGATTTAAACCTTAAGTA 1226
|||||
1899 CATCATTCCTTCTAGTGTAGCTCCAGCTGCTGTAGACCTAGATTTAAACCTTAAGTA 1958
|||||
1227 AGATGCTGGGTATAGAACGCTAAGAAATTTCCCCCAAGGACTCTTGCTTCTCTTAAGCCC 1286
|||||
1959 AGATGCTGGGTATAGAACGCTAAGAAATTTCCCCCAAGGACTCTTGCTTCTCTTAAGCCC 2018
|||||
1287 TTTCTGGCTTCGTTTATGGTCTTCAATTTAAAGTATAAGCCTTAACCTTGTGCTAGTCTTAA 1346
|||||
2019 TTTCTGGCTTCGTTTATGGTCTTCAATTTAAAGTATAAGCCTTAACCTTGTGCTAGTCTTAA 2078
|||||
1347 GGAGAAACCTTTACCAACAAGTTTATCATTTGAAGACATATTTGAACACCCCTTAT 1406
|||||
2079 GGAGAAACCTTTACCAACAAGTTTATCATTTGAAGACATATTTGAACACCCCTTAT 2138
|||||
1407 TTGTGGGATTTGAGAAGGGTGAATAGAGGCTTTCCCTTTGTTGTTAGGACTT 1466
|||||

Db 2139 TTGTGGGATTGAGAAGGGTGAATAGAGGCTTGAGACTTTCTTTTGTGTGTAGGACTT 2198
Qy 1467 GGAGAGAAATCCCTCGACTTTCTACTAACCTCTGACATACTCCCCACACCCAGTTGAT 1526
|||||
Db 2199 GGAGAGAAATCCCTCGACTTTCTACTAACCTCTGACATACTCCCCACACCCAGTTGAT 2258
Qy 1527 GCCTTTCCGTAATAAAAAAGATTGGGATT 1554
|||||
Db 2259 GCCTTTCCGTAATAAAAAAGATTGGGATT 2286
|||||
RESULT 12
ABAL5532
ID ABAL5532 standard; DNA; 3369 BP.
XX
AC ABA15532;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 7863.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.


```
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234937.
PR 25-SEP-2000; 2000US-0234938.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
XX
XX Disclosure; SEQ ID NO 7863; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
XX (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3369 BP; 850 A; 724 C; 788 G; 1007 T; 0 other;
XX
XX Query Match 55.8%; Score 868; DB 22; Length 3369;
XX Best Local Similarity 100.0%; Pred. No. 3.5e-272;
XX Matches 868; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 687 AGGAGAAATGTGATCCGAGAAATTAAGTATACCTTAATAGCTATACCGGGGCCAAGAACTAT 746
Db 2491 AGGAGAAATGTGATCCGAGAAATTAAGTATACCTTAATAGCTATACCGGGGCCAAGAACTAT 2550
QY 747 CAAAGCTGGAGACAAATATCCCTGAGGACGCCCTGTGCTTCAACCCGCCACAGTGT 806
Db 2551 CAAAGCTGGAGACAAATATCCCTGAGGACGCCCTGTGCTTCAACCCGCCACAGTGT 2610
```


XX	29-SEP-2000;	2000US-0236327.	XX	WPI; 2001-541565/60.
DR	29-SEP-2000;	2000US-0236367.	XX	Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PR	29-SEP-2000;	2000US-0236368.	PT	useful for preventing, diagnosing and/or treating nervous system
PR	29-SEP-2000;	2000US-0236369.	PT	cancers and metastases
PR	29-SEP-2000;	2000US-0236370.	XX	Disclosure; SEQ ID NO 7864; 1701pp + Sequence Listing; English.
PR	02-OCT-2000;	2000US-0236802.	PS	
PR	02-OCT-2000;	2000US-0237037.	XX	
PR	02-OCT-2000;	2000US-0237038.	XX	The invention relates to novel genes (ABA11004-ABA21534) and proteins
PR	02-OCT-2000;	2000US-0237039.	CC	(ABBI4678-ABBI8001) useful for preventing, treating or ameliorating
PR	02-OCT-2000;	2000US-0237040.	CC	medical conditions e.g. by protein or gene therapy. The genes are
PR	13-OCT-2000;	2000US-0239935.	CC	isolated from a range of human tissues disclosed in the specification.
PR	13-OCT-2000;	2000US-0239937.	CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful
PR	20-OCT-2000;	2000US-0241785.	CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
PR	20-OCT-2000;	2000US-0241786.	CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone
PR	20-OCT-2000;	2000US-0241787.	CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
PR	20-OCT-2000;	2000US-0241808.	CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune
PR	20-OCT-2000;	2000US-0241809.	CC	hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
PR	20-OCT-2000;	2000US-0241826.	CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative
PR	01-NOV-2000;	2000US-0246617.	CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;
PR	01-NOV-2000;	2000US-0246617.	CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
PR	08-NOV-2000;	2000US-0246475.	CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
PR	08-NOV-2000;	2000US-0246475.	CC	and parasitic infections.
PR	08-NOV-2000;	2000US-0246476.	CC	Note: The sequence data for this patent did not form part of the
PR	08-NOV-2000;	2000US-0246477.	CC	printed specification, but was obtained in electronic format directly
PR	08-NOV-2000;	2000US-0246478.	CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
PR	08-NOV-2000;	2000US-0246523.	XX	
PR	08-NOV-2000;	2000US-0246524.	SQ	Sequence 3369 BP; 849 A; 724 C; 789 G; 1007 T; 0 other;
PR	08-NOV-2000;	2000US-0246525.		
PR	08-NOV-2000;	2000US-0246526.		
PR	08-NOV-2000;	2000US-0246527.		
PR	08-NOV-2000;	2000US-0246528.		
PR	08-NOV-2000;	2000US-0246532.		
PR	08-NOV-2000;	2000US-0246609.		
PR	08-NOV-2000;	2000US-0246610.		
PR	08-NOV-2000;	2000US-0246611.		
PR	17-NOV-2000;	2000US-0246613.		
PR	17-NOV-2000;	2000US-0249207.		
PR	17-NOV-2000;	2000US-0249208.		
PR	17-NOV-2000;	2000US-0249209.		
PR	17-NOV-2000;	2000US-0249210.		
PR	17-NOV-2000;	2000US-0249211.		
PR	17-NOV-2000;	2000US-0249212.		
PR	17-NOV-2000;	2000US-0249213.		
PR	17-NOV-2000;	2000US-0249214.		
PR	17-NOV-2000;	2000US-0249215.		
PR	17-NOV-2000;	2000US-0249216.		
PR	17-NOV-2000;	2000US-0249217.		
PR	17-NOV-2000;	2000US-0249218.		
PR	17-NOV-2000;	2000US-0249244.		
PR	17-NOV-2000;	2000US-0249245.		
PR	17-NOV-2000;	2000US-0249264.		
PR	17-NOV-2000;	2000US-0249265.		
PR	17-NOV-2000;	2000US-0249297.		
PR	17-NOV-2000;	2000US-0249299.		
PR	17-NOV-2000;	2000US-0249300.		
PR	01-DEC-2000;	2000US-0250391.		
PR	01-DEC-2000;	2000US-0251160.		
PR	05-DEC-2000;	2000US-0251030.		
PR	05-DEC-2000;	2000US-0251988.		
PR	05-DEC-2000;	2000US-0256719.		
PR	06-DEC-2000;	2000US-0251479.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251868.		
PR	08-DEC-2000;	2000US-0251869.		
PR	08-DEC-2000;	2000US-0251989.		
PR	08-DEC-2000;	2000US-0251990.		
PR	11-DEC-2000;	2000US-0254097.		
PR	05-JAN-2001;	2001US-0259678.		
XX	(HUMA -) HUMAN GENOME SCI INC.			
PA	Rosen CA, Barash SC, Ruben SM;			

QY 1287 TTCTGGCTTCGTTTATGCTCTTCATTAAGGCTATAGCCTAACTTTGCTAGTCCCTAA 1346
 Db 3091 TTCTGGCTTCGTTTATGCTCTTCATTAAGGCTATAGCCTAACTTTGCTAGTCCCTAA 3150
 QY 1347 GGAGAAACCTTTAACCACAAAGTCTTTATCATTTGAAGACAATATTGAACAACCCCTATT 1406
 Db 3151 GGAGAAACCTTTAACCACAAAGTCTTTATCATTTGAAGACAATATTGAACAACCCCTATT 3210
 QY 1407 TTGTGGGATGAGAGGGGTAATAGAGGCTTGAGACTTCTCTTTGTGTAGGACTTT 1466
 Db 3211 TTGTGGGATGAGAGGGGTAATAGAGGCTTGAGACTTCTCTTTGTGTAGGACTTT 3270
 QY 1467 GGAGAGAAATCCCTGACATTTCTACTAACCTCTGACATACCTCCACACCCAGTTGAT 1526
 Db 3271 GGAGAGAAATCCCTGACATTTCTACTAACCTCTGACATACCTCCACACCCAGTTGAT 3330
 QY 1527 GGCCTTCGTAATAAAAGATTGGGATT 1554
 Db 3331 GGCCTTCGTAATAAAAGATTGGGATT 3358

RESULT 14
 \15981

AAA15981 standard; cDNA; 774 BP.

XX AC AAAL5981;

XX DT 12-JUN-2000 (first entry)

XX DE Human protein clone HP10392 coding sequence.

XX KW Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.

XX OS Homo sapiens.

XX PN WO200005367-A2.

XX PD 03-FEB-2000.

XX PF 22-JUL-1999; 99WO-JP03929.

XX 24-JUL-1998; 98JP-0208820.

XX 07-AUG-1998; 98JP-0224105.

XX 25-AUG-1998; 98JP-0238116.

XX 09-SEP-1998; 98JP-0254736.

XX 29-SEP-1998; 98JP-0275505.

XX (SAGA) SAGAMI CHEM RES CENT.

XX (PROT-) PROTEGENE INC.

XX Kato S, Kimura T;

XX WPI; 2000-182694/16.

XX P-PSDB; AAY94893.

XX Novel human proteins having hydrophobic domains useful for treating
 XX osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 XX multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke

XX Claim 3; Page 328; 351pp; English.

XX This sequence encodes a human protein of the invention, which has

CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.

XX SQ Sequence 774 BP; 200 A; 200 C; 190 G; 184 T; 0 other;

Query Match 49.8%; Score 774; DB 21; Length 774;

Best Local Similarity 100.0%; Pred. No. 7.5e-242;

Matches 774; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ATGCGGGTTCGGACCTTAATGCTCTCGTGTATTCGGTCCCGGACATTTACGATGG 118

Db 1 ATGCGGGTTCGGACCTTAATGCTCTCGTGTATTCGGTCCCGGACATTTACGATGG 60

QY 119 CTCGCCCCACCTTACTACCTTCTGCGCCCTGCTCTCTGCTTCTTCTACTCTGAGG 178

Db 61 CTCGCCCCACCTTACTACCTTCTGCGCCCTGCTCTCTGCTTCTTCTACTCTGAGG 120

QY 179 AAACCTGCGCGCGCTCTGCGCACGGTCTGCCACCCACGGAAGAGCGGTAAACCGGTGAC 238

Db 121 AAACCTGCGCGCGCTCTGCGCACGGTCTGCCACCCACGGAAGAGCGGTAAACCGGTGAC 180

QY 239 TTTGACTGGAGAGAAAGTGAGATCCCTGATGTTTCTCAGTGCCTATGTTGATGAAGAAC 298

Db 181 TTTGACTGGAGAGAAAGTGAGATCCCTGATGTTTCTCAGTGCCTATGTTGATGAAGAAC 240

QY 299 CGCAGATCCATGTTCTCTGATGACGTGCAAAACCCCTATATATGCGCCCTCAGTATATC 358

Db 241 CGCAGATCCATGTTCTCTGATGACGTGCAAAACCCCTATATATGCGCCCTCAGTATATC 300

QY 359 AAGTACTTCAATGATAAAACCACTTATGAGGAACTAGAACGGGACAAAGAGGTCTACTGG 418

Db 301 AAGTACTTCAATGATAAAACCACTTATGAGGAACTAGAACGGGACAAAGAGGTCTACTGG 360

QY 419 ATTGTGGAGTTCTTTGCCAATTTGGTCTAATGACTGCCAATCATTTGCCCTTCTATGCT 478

Db 361 ATTGTGGAGTTCTTTGCCAATTTGGTCTAATGACTGCCAATCATTTGCCCTTCTATGCT 420

QY 479 GACCTCTCCCTTAATAACACTGTACAGGGCTAAATTTTGGGAAGTGTGATTTGGAGCC 538

Db 421 GACCTCTCCCTTAATAACACTGTACAGGGCTAAATTTTGGGAAGTGTGATTTGGAGCC 480

QY 539 TATACTGATGTTAGTACGGGTACAAAGTACACATCACCCCTCACCAGCACTCCCT 598

Db 481 TATACTGATGTTAGTACCGGTACAAAGTGAGCATACACCCCTCACCAAGCAACTCCCT 540
Qy 599 ACCCTGATCCCTGTTCCCAAGGTGGCAAGGAGCAATGCGCGGCCACAGATTGACAAAGAA 558
Db 541 ACCCTGATCCCTGTTCCCAAGGTGGCAAGGAGCAATGCGCGGCCACAGATTGACAAAGAA 600
Qy 659 GGACGGGCTGTCATGACCTTCTCTGAGGAGAAATGATCCGAGAAATTTAACTTAAAT 718
Db 601 GGACGGGCTGTCATGACCTTCTCTGAGGAGAAATGATCCGAGAAATTTAACTTAAAT 560
Qy 719 GAGCTATACCGCGGCCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAG 778
Db 661 GAGCTATACCGCGGCCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAG 720
Qy 779 CCTGTGGCTTCAACCCCAACACAGTGTGAGATGGGGGAAACAAGAAGATAAA 832
721 CCTGTGGCTTCAACCCCAACACAGTGTGAGATGGGGGAAACAAGAAGATAAA 774

RESULT 15
AAS83076 standard; cDNA; 2030 BP.
XX AAS83076;
AC AAS83076;
X
Dt 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #18880.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
X
X Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB: ABG18889.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 18880; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2030 BP; 500 A; 476 C; 516 G; 523 T; 15 other;
Query Match 39.4%; Score 613; DB 23; Length 2030;
Best Local Similarity 78.7%; Pred. No. 5; se-189;
Matches 1403; Conservative 0; Mismatches 150; Indels 230; Gaps 47;
Qy 2 GGGGAGCGGGCGGAGACCTACGACCGCGGAGAGTGGCCCTTACGGCCGAAAGATG 61
Db 44 GGGGAGCGGGCGGAGACCTACGACCGCGGAGAGTGGCCCTTACGGCCGAAAGATG 103
Qy 62 GGGGCTTTGGCACCTCTAATT--GCTCTCTGTTATTCGGTGGCGCGAC--TTTCACGATGG 118
Db 104 GCGGCTTTGGCACCTCTAATTGCTCTCGTGTATTTCGGTGGCGGACATTTTCACGATGG 163
Qy 119 CTCGCCCAACCTTACTACCTTCTGTCGG--CCCTGCTCTCTGTCGCTT--CCTACTCTGA 176
Db 164 CTCGCCCAACCTTACTACCTTCTGTCGGCCCTCTCTCTGCTGCTTCCCTACTCTGA 223
Qy 177 GGAACCTGCCGCGCTCTGCCAGGTCTGCCACCCAA--CGCGAAGACGGTA--CCCGTG 234
Db 224 GGAACCTGCCGCGCTCTGCCAGGTCTGCCACCCAA--CGCGAAGACGGTA--CCCGTG 283
Qy 235 TGACTTTGACTGGAGAG--AAGTGGAGATCTCTGATGTTTCTCAGTGCATTTGATGATGA 293
Db 284 TGACTTTGACTGGAGAGAAAGTGGAGATCTCTGATGTTTCTCAGTGCATTTGATGATGA 343
Qy 294 AGAA--CCGCAGATCCA----- 308
Db 344 AGAACCAGATGATCCATCTCTGGAGCAACATATAGCAACATTTTTCATGTTTAAAGTAA 403
Qy 309 ----- 308
Db 404 AAGTGGCCAAACAATTTCTTTCTCCCGCTTGGATATTGCGATGGCGCTACTTTACATC 463
Qy 309 -----TGTTCTCTGATGAGTGCACAAACCC-----CCCTATATATGGCCCTG 350
Db 464 ACACCTCTGATAGTGTCTCTGATGAGTGCACAAACCCCGCTCTCCCATATATGGCCCTG 523
Qy 351 AGTATATCAAGTACTTCAAT--GATAAACCATTGATGAGAACTAGAACGGGACAAAGAG 409
Db 524 AGTATATCAAGTACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 583
Qy 410 GTCACCTGGA--TTGTGGAGTCTTTGCCAATTTGGTCTTAATGACTGCCAATCAATTTGCC 467
Db 584 GTCAATTTGGAATTTGTTGGAGTCTTTTGGCAATTTGGTCTTAATGACTGCCAATCAATTTGCC 643
Qy 468 CTATCTATCTGACCTCTCCCTTAAATACAACTGTACAGGCTTAAATTTTGGGAAGTGG 527
Db 644 CTATCTATCTGACCTCTCCCTTAAATACAACTGTACAGGCTTAAATTTTGGGAAGTGG 703
Qy 528 ATGTTGGAGCTATCTGATGTTAGTACGCGGTACAAAGTGGAGACATCACCCCTCACCA 587
Db 704 ATGTTGGAGCTATCTGATGTTAGTACGCGGTACAAAGTGGAGACATCACCCCTCACCA 763
Qy 588 A--GCAACTCCCTTACCTT--GATCCTGTTTCAAGTGGCAA--GGAGGCAATGGCGGCGCC 642
Db 764 AGGCAACTCCCTTACCTTGGATCTGTTTCCAAAGTGGGCAAGGAGGCAATTCGCGGCGGCC 823
Qy 643 ACAGATT-----GACAAGAAAGGCGGCTGCTCTCATGGACCTTCTCTG----- 686
Db 824 ACAGATTGCAAGAAAGGAGCGGCGTGTGTTTTCATGGGACCTTCTCTGAGTACCTGA 883
Qy 687 -----AGGAGAAATGTCGAGAAATTAACCTTAAATGAGCTATACCGGCGGCC 736
Db 884 AAGGAGGCGAGAGATGTCGAGAAATTAACCTTAAATGAGCTATACCGGCGGCC 943

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 12, 2003, 09:47:42 ; Search time 185 Seconds
(without alignments)
3140.625 Million cell updates/sec

Title: US-09-954-846-2
Perfect score: 1359
Sequence: 1 MAVIAPLIALVYSPRLSRW.....EQPVASTPTTSDGENKKDK 258

Scoring table:
BLOSUM62 Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_SCOOL/US09954846/runat_09072003.102144.12148/app_query.fasta.1.455
-DB=N_Geneseq_101002 -QFAST=fastap -SUFFIX=ing -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09954846.ecgn.1.1.200.0/runat_09072003.102144.12148 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPHLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tabase : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1359	100.0	774	21	AAAL5981 Human protein clon
2	1359	100.0	1527	21	AAAL5991 Human protein clon
3	1359	100.0	1598	21	AAFL16316 Human prostate can
4	1326	97.6	1618	20	AAFL16316 Protein PRO270 CDN
5	1326	97.6	1620	22	AAAS45940 Human DNA encoding
6	1326	97.6	1730	22	ABAO9248 Human CGI-31 prote
7	1322	97.3	1699	19	AAV34294 Human secreted pro
8	1313	96.6	1621	22	AAV72409 Human secreted pro
9	1313	96.6	1652	19	AAV34316 Human secreted pro
10	1308	96.2	1659	24	ABL49614 Prostaglandin E2 (
11	1299	95.6	1631	22	AAFL16316 Human cDNA encodin
12	994	73.1	774	21	AA242527 Human 5' EST isola
13	956	70.3	726	22	AAFL16316 Primer specific fo
14	925.5	68.1	2030	23	AAFL16316 DNA encoding novel
15	622	45.8	439	21	AAAL1964 Human secreted exp
16	620	45.6	461	21	AAAL1964 Human secreted pro
17	579.5	42.6	3369	22	ABAL15533 Human nervous syst
18	579.5	42.6	3369	22	ABAL15533 Human nervous syst
19	514	37.8	444	20	AAAL1205 Human secreted pro
20	498	36.6	2297	22	ABAL15531 Human nervous syst
21	476.5	35.1	919	23	ABLI17145 Drosophila melanog
22	476.5	35.1	2919	23	ABLI17145 Drosophila melanog
23	447	32.9	430	20	AAAL1202 Human secreted pro
24	431.5	31.8	5423	23	AAFL16316 DNA encoding novel
25	431.5	31.8	5423	24	ABL68483 Kidney cancer rela
26	430	31.6	452	20	AAAL1203 Human secreted pro
27	424	31.2	307	23	AAFL16316 DNA encoding novel
28	422	31.1	469	21	AAAL1203 Human 5' EST isola
29	415	30.5	307	24	ABK45061 CDNA encoding colo
30	372	27.4	274	21	AAAL1585 Human secreted exp
31	353.5	26.0	756	23	ABV21932 Human prostate exp
32	353.5	26.0	756	23	ABV27764 Human prostate exp
33	336	24.7	1895	23	AAFL16316 DNA encoding novel
34	323	23.8	1409	22	ABAL15530 Human nervous syst
35	280	20.6	531	21	AAAL1202 Human 5' EST isola
36	269	19.8	153	22	AAAL19490 Human breast cance
37	269	19.8	433	22	AAAL09103 Human breast cance
38	232	17.1	370	23	ABV02989 Human prostate exp
39	175	12.9	367	22	ABAL1541 Human nervous syst
40	163.5	12.0	375	23	ABV33303 Human prostate exp
41	163.5	12.0	375	23	ABV42226 Human prostate exp
42	163.5	12.0	389	23	ABV12158 Human prostate exp
43	130.5	9.6	3550	9	AAAL1540 Sequence of pTrx-2
44	130.5	9.6	3550	13	AAQ31623 pTrx-2 containing
45	129.5	9.5	3552	11	AAQ05875 pTrx-2 plasmid seq

ALIGNMENTS

RESULT 1
AAAL5981
ID AAAL5981 standard; cDNA; 774 BP.
XX
AC AAAL5981;
XX
XX
XX 12-JUN-2000 (first entry)
DT
DE Human protein clone HPI0392 coding sequence.

XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;

KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.

XX Homo sapiens.

XX WO200005367-A2.

XX 03-FEB-2000.

XX 22-JUL-1999; 99WO-JP03929.

XX 24-JUL-1998; 98JP-0208820.

XX 07-AUG-1998; 98JP-0224105.

XX 25-AUG-1998; 98JP-0238116.

XX 09-SEP-1998; 98JP-0254736.

XX 29-SEP-1998; 98JP-0275505.

XX (SAGA) SAGAMI CHEM RES CENT.

XX (PROT-) PROTEGENE INC.

XX Kato S, Kimura T;

XX WPI; 2000-182694/16.

XX P-PSDB; AAY94893.

XX Novel human proteins having hydrophobic domains useful for treating

XX osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,

XX multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -

XX Claim 3; Page 328; 351pp; English.

XX This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.

XX Sequence 774 BP; 200 A; 200 C; 190 G; 184 T; 0 other;
 XX Alignment Scores:
 XX Pred. No.: 3,41e-155 Length: 774
 XX Score: 1359.00 Matches: 258
 XX Percent Similarity: 100.00 Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-954-846-2 (1-258) x AAA15981 (1-774)

QY 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTTP 20
 DB 1 ATGGCGGCTTGGCACCTCTAATGTCTCGGTGATTCGGTCCGGACATTTCCACGATGG 60
 QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
 DB 61 CTGCGCCAAACCTTACTACCTTCTGTGCGCCCTCTCTCTGCTGCTTCTTCTACTCTGAGG 120
 QY 41 LysLeuProLeuLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 DB 121 AAACCTGCCCGCGCTCTGCCACGCTCTGCCACCGCTCTGCCACCGCTCTGCCACCGCTCTGCC 180
 QY 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsp 80
 DB 181 TTTGACTGGAGAGAGAGTGGAGATCCCTGATGTTTCTCAGTGCCTTGTGATGATGATGATGATG 240
 QY 81 ArgArgSerMetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIle 100
 DB 241 CGCAGATCCATGTCTCTGATGACGTGCAAAACCCCTATATATGGCCCTGAGTATATC 300
 QY 101 LysTyrPheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrp 120
 DB 301 AAGTACTTCAATGATATAAACCAATTTGATGAGGAAGTAAAGCGGACGAGAGGGTCACTTGG 360
 QY 121 IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAla 140
 DB 361 ATGTGGAGTCTTTCGCCAATTTGGTCTAATGACTGCCAATTCATTTGCCCTTCTATGCT 420
 QY 141 AspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArg 160
 DB 421 GACCTCTCCCTTAATACAACTGTACAGGCTAAATTTGGGAAGGTGGATGTTGGACGC 480
 QY 161 TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro 180
 DB 481 TATACTGATGTTAGTACGGGTACAAAGTACAGCATCACCCCTCACCAAGCACTCCCT 540
 QY 181 ThrLeuIleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
 DB 541 ACCTGATCCTGTTCGAAGGTGGCAAGGAGGCAATGCGGCGGCACAGATGACAAAGAA 600
 QY 201 GlyArgAlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn 220
 DB 601 GGACGGGCTGTCTCATGGACCTCTCTGAGGAGATGTGATCCGAGAAATTAACCTTAAT 660
 QY 221 GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGln 240
 DB 661 GAGCTATACAGCGGCGCCAAAGAACTATCAAAAGCTGGAGACAATATCCCTGAGGAGCAG 720
 QY 241 ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
 DB 721 CCTGTGGCTTCAACCCCGCCACACAGTGTGAGATGGGGAACAAAGAGGATATAA 774

RESULT 2

AAA15991

ID AAA15991 standard; cDNA; 1527 BP.

XX AAA15991;

AC AAA15991;

XX 12-JUN-2000 (first entry)

XX Human protein clone HP10392 full length coding sequence.

XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;

XX cytokine production; cell proliferation; cell differentiation;

XX immune deficiency; infectious disease; autoimmune disorder; asthma;

XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;

XX allergic reaction; osteoporosis; osteoarthritis; periodontal disease;

Db 543 TCCCTTAATCAACTGACAGGGCTAAATTTTGGGAAGGTGGATGTTGGACGCTATACT 602
 Qy 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
 Db 603 GATGTTAGTACGCGGTACAAAGTGAGACATCACCCCTCACCAAGCAACTCCCTACCCCTG 662
 Qy 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
 Db 663 ATCTCTGTTCCAAAGGTGGCAAGAGGCAATGCGGCGCCACAGATTGACAAGAAAGGACGG 722
 Qy 203 AlaValSerTTP-ThrPheSerGluGluAsnValIleArgGluPheAsnLeuAspGluLe 222
 Db 723 CTGTCTCATGGACCTTCTCTGAGGAGAAAGTGATCCGAGAAATTAATTAATGAGCT 782
 Qy 222 uTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVa 242
 Db 783 ATACAGCGGGCCCAAGAAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGT 842
 Qy 242 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
 Db 843 GGCTTCAACCCCAACACAGTGTCAAGTGGGGAACAAGAAAGGATAAA 891
 JLT 9
 ID AAV34316 standard; DNA; 1652 BP.
 XX AAV34316
 AC AAV34316;
 XX 29-JAN-1999 (first entry)
 DE Human secreted protein gene 9 clone HJBCD89.
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX Homo sapiens.
 OS
 XX WO9840483-A2.
 PN
 XX 17-SEP-1998.
 PD
 XX 12-MAR-1998; 98WO-US04858.
 XX
 XX 19-DEC-1997; 97US-0068368.
 XX 14-MAR-1997; 97US-0040710.
 PR 14-MAR-1997; 97US-0040762.
 PR 30-MAY-1997; 97US-0048100.
 PR 30-MAY-1997; 97US-0048189.
 PR 30-MAY-1997; 97US-0048357.
 PR 30-MAY-1997; 97US-0050934.
 PR 06-JUN-1997; 97US-0048970.
 PR 05-SEP-1997; 97US-0057765.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Ferrie AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;
 PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;
 PI Wei YF, Young PE, Zeng Z;
 XX WPI: 1998-520811/44.
 DR P-PSDB; AAW5226.
 DR
 XX Isolated human poly:nucleotide(s) encoding secretory peptide(s) -
 PT used to develop products for the diagnosis and treatment of e.g.
 PT inflammation, cancers, CNS disorders or immune system disorders
 XX

PS Claim 1; Page 143-144; 201pp; English.
 XX This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number, and the clone it is derived
 CC from, are detailed in the descriptor line. The gene can be used to
 CC generate fusion proteins by linking to the gene to a human immunoglobulin
 CC Fc portion (e.g. AAV34277) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 28 novel genes and their fragments (nucleic
 CC acid sequences: AAV34286-V34325; amino acid sequences AAW5196-W5235)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 28
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV34286 for described uses).
 XX
 SQ Sequence 1652 BP; 423 A; 396 C; 379 G; 448 T; 6 other;
 Alignment Scores:
 Pred. No.: 4,06e-149 Length: 1652
 Score: 1313.00 Matches: 254
 Percent Similarity: 87.03% Conservative: 1
 Best Local Similarity: 86.69% Mismatches: 0
 Query Match: 96.62% Indels: 38
 DB: 19 Gaps: 1
 US-09-954-846-2 (1-258) x AAV34316 (1-1652)
 Qy 4 LeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrpLeuAlaGln 23
 Db 1 TTGGCACCCTCTAATTGCTCTGCTGATTCGGTGGCGGACTTTCAGATGGCTCGCCCAA 60
 Qy 24 ProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArgLysLeuPro 43
 Db 61 CCTTACTACTCTCTGCGCCCTGCTCTGCTGCTCTCTCTACTCTGAGGAACCTGCGG 120
 Qy 44 ProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAspPheAspTrp 63
 Db 121 CGGCTCTGCGACGGTCTGCCACCCCAACGGAAGACGGTAACCCGTGCTGACTTGTGCTG 180
 Qy 64 ArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsnArgArgSer 83
 Db 181 AGAGAAGTGAGATCCTGATGTTTCTCAGTGCCATTGTGATGATGAAGAACCCGACATCC 240
 Qy 83 ----- 83
 Db 241 ATCACTGTGGACCAACATATAGGCAACATTTTCATGTTTAGTAAGTGGCCCAACAATT 300
 Qy 84 -----MetPhe 85
 Db 301 CTTTCTTCCGCTTGGATATTGCGATTCGCGCTACTTTACATCACACTCTGCAATGTTTC 360
 Qy 86 LeuMetThrCysLysProLeuTyrMetGlyProGluTyrIleLysTyrPheAsnAsp 105
 Db 361 CTGATGACGTGCAAAACCCCTATATATGCGGSCCTGAGTATATCAAGTACTCAATGAT 420
 Qy 106 LysThrIleAspGluLeuGluArgAspLysArgValThrTrpIleValGluPhePhe 125
 Db 421 AAACACATTGATCAGGAACACTAGACGGGACAGAGGGTCACCTGGATTGGAGTTCTTT 480
 Qy 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
 Db 481 GCCAATTGCTCTAATGACTGCCCAATCATTTGCCCTATCTATGCTGACCTCTCCCTTAA 540
 Qy 146 TyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThrAspValSer 165
 Db 541 TACAACCTGTACAGGGCTAAATTTTGGAGGTTGATGTTGGACGCTATACTGATGTACT 600
 Qy 166 ThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPhe 185
 Db 601 ACGCGGTACAAAGTGAGCACATCACCCTTCACCAAGCAACTCCCTACCTGATGCTGTTTC 660

186	GlnGlyGlyLysGluAlaMetArgProGlnIleAspLysLysGlyArgAlaValSer	205
Qy		
661	CAAGGTGGCAAGGAGGCAATCGCGGGCCACAGATTGACAGAAAAGACGGGCTGTCTCA	720
Db		
206	TrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArg	225
Qy		
721	TGGACCTTCTCTGAGGAGAAATGTATCCGAGAATTAACTAAATGAGCTATACCAGCGG	780
Db		
226	AlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThr	245
Qy		
781	GCCAAAGAACTATCAAGGCTGGAGACATAATCCCTGAGGAGCACCGCTGGGCTTCACCC	840
Db		
246	ProThrThrValSerAspGlyGluAsnLysLysAspLys	258
Qy		
841	CCACCAACAGTGTCCAGATGGGGAACAACAAAGAAGGATAAA	879
Db		

.ULT 10

..5L49614

ID ABL49614 standard; DNA; 1659 BP.

AC ABL49614:

01-JUN-2002 (first entry)

Prostaglandin E₂ (PGE₂) en

XX
KW Prostaglandin E2: stomach cancer: gene: ds.

Unidentified

XX	Key	Location/Qualifiers
FH		

FT	CDS	31.1.921
FT	CDS	31.1.921

$$\frac{F_T}{*taq} = a$$

```
FT /product="prostaglandin E2"
```

PN KR2001081233-A.

PD 29-AUG-2001.

11-FEB-2000: 2000KR-0006416.

11-FEB-2000: 2000KR-0006416-XX
PRXX
--
-- (NAHE-) NAT INST HEALTH IN KOREA

Kim GC. Nam M.T. Park MS.

XX
DP
WPT: 2002-136917/18

DR WPI; 2002-136917-
DR P-PSDB: ABB06300

AA	Novel gene expressed by prostaglandin E2 (PGE2) in stomach cancer cell
PT	line snu-1 and its amino acid sequence -
PT	Novel gene expressed by prostaglandin E2 (PGE2) in stomach cancer cell
AA	line snu-1 and its amino acid sequence -

XX
 XXXX

PS Disclosure; Page. 12; 12pp; Korean.

The present sequence encodes prostaglandin E2 (PGE2) which is expressed in the stomach cancer cell line SNU-1. The PGE2 gene can be used in the treatment of stomach cancer.

XX
TO CHURCHES

SQ Sequence 1659 BP; 426 A; 404 C; 380 G; 448 T; 1 other;

Alignment Scores:

Alignment Scores:

Argument Scores:	
Pred. No.:	1.65e-148
Length:	1659

Score:	1308.00	Matches:	255
--------	---------	----------	-----

Percent Similarity: 86.49%

Best Local Similarity: 86.15%
Mismatches: 2

Query Match:	96.25%	Indels:	38
--------------	--------	---------	----

DB:	24	Gaps:	1
-----	----	-------	---

US-09-954-846-2 (1-258) X ABL49614 (1-1659)

Qy 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArqLeuSerArqTrp 20

Db	31	ATGCGCGTCTGGCACCTCTAAATTCCTCTCTGCTGATTTCGGTGGCGCGACTTTTCAGATGG	90
Qy	21	LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg	40
Db	91	CTCGCCCAACCTTACTACCTTCTCTCGCGCCCTGCTCTCTGCTGCTCTCTACTCGTGAGG	150
Qy	41	LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp	60
Db	151	AAACTGCGCGCGTCTCTGCCACGGTCTGCCACCGCAACCAACGCAAGACGGTAACCGGTGTGAC	210
Qy	61	PheAspTrpArgGluValGluIleuMetPheLeuSerAlaIleValMetMetLysAsn	80
Db	211	TTTGACTGGAGAGAAGTGGAGATCCTGATGTTTCTCAGTGCAATGTGATGATGAAGAAC	270
Qy	81	ArgArgSer-----	83
Db	271	CGCAGATCCACTACTGTGGAGCAACATATAGGCCAACATTTTCATGTTTAGTAAAGTGGCC	330
Qy	83	-----	83
Db	331	AACACAATCTTTTCTTCGCGCTTGGATATTCGCATGGCGCTACTTTACATCACACTCTGC	390
Qy	84	---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr	102
Db	391	ATAGTGTCTCTGATGAGCTGCACACCCCTCATATATGGCCCTGAGTATATCAAGTAC	450
Qy	103	PheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleVal	122
Db	451	TTCAATGATAAACCATGTATGAGAACTAGAACGGGCAAGAGGGTCACTTGGATTGTG	510
Qy	123	GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu	142
Db	511	GAGTTCCTTGCCAAATTGGTCTAATGACTGCCAATCATTTGCCCTATCTATGCTGACCTC	570
Qy	143	SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr	162
Db	571	TCCCTTAAATACAACTGTACAGGCGCTAAATTTTGGAGGTGGATGTTGGACGCTATACT	630
Qy	163	AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu	182
Db	631	GATGTTAGTACGGGTACAAAGCAGACATCACCCCTCACCAAGCAACTCCCTACCCCTG	690
Qy	183	IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg	202
Db	691	ATCCTGTTCCAAGGTGCGCAAGGAGCAATGCGCGCGCCACAGATTGACAAGAAGGACGG	750
Qy	203	AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu	222
Db	751	GCTGTCFCAGGGACCTTCTCTGAGGAGAAATGTGATCGAGAAATTAACCTTAATGAGCTA	810
Qy	223	TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal	242
Db	811	TACCAGGGGCGCAAGAAACATATCAAGAGCTGGAGACAATATCCCTGAGGAGCAGCCTGTG	870
Qy	243	AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys	258
Db	871	GCITCAACCCCCACACAGTGTGATGGGGAACCAACAGAGGATATA	918
RESULT	11		
AAF93762			
ID	AAF93762	standard; cdna; 1631 BP.	
XX	AAF93762;		
XX			
DT	23-MAY-2001	(first entry)	
XX			
DE		Human cdna encoding a membrane or secretory protein clone P5EC0045.	
XX			
KW		Human; secretory protein; membrane protein; vaccine; gene therapy;	
KW		rheumatoid arthritis; diabetes; ss.	
XX			
OS		Homo sapiens.	

XX EP1067182-A2.
 XX 10-JAN-2001.
 XX 07-JUL-2000; 2000EP-0114090.
 XX 08-JUL-1999; 93JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Kawal Y, Sugiyama T, Hayashi K;
 PI P-PSDB; AAB88335.
 XX WPI; 2001-093989/11.
 DR P-PSDB; AAB88335.
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 PT Claim 1; SEQ ID 37; 609pp + CD ROM; English.
 PS This invention relates to nucleic acid sequences AAF93744 - AAF93916
 vv which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent-assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 XX
 SQ Sequence 1631 BP; 411 A; 397 C; 378 G; 445 T; 0 other;

Alignment Scores:
 d. No.: 2e-147 Length: 1631
 re: 1299.00 Matches: 256
 Percent Similarity: 86.82% Conservative: 1
 Best Local Similarity: 86.49% Mismatches: 1
 Query Match: 95.58% Indels: 39
 DB: 22 Gaps: 1

US-09-954-846-2 (1-258) x AAF93762 (1-1631)

QY 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTTP 20
 Db 16 ATGGCGGTCTGGACCTTAATGCTCGGTATTCGGTATTCGGCGGACTTCCACGATG 75
 QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
 Db 76 CTCGCCAACCTTACTACCTTCTGTGGCCCTGCTCTGCTGCTTCTACTCGTGAGG 135
 QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 Db 136 AAATCGCCCGCTCTGCGACGCTGCTGCCACCCACGACGAGACGTAACCCGCTGTAC 195
 QY 61 PheAspTTPArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80

Db 196 TTTCAGTGGAGAGAGAGAGATCCCTGATGTTTCTCAGTCCCATTTGTGATGATGAAGAAC 255
 QY 81 ArgArgSer-----
 Db 256 CGCAGATCCATCAGTCTGGAGCAACATATAGGCAACATTTTTCATGTTTAAAGTGCC 315
 QY 83 -----
 Db 316 AACACAATTCTTTCTCCGCTTGGATATTCGCATGGGCTACTTTTACATCAGCACTGCG 375
 QY 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
 Db 376 ATAGTCTTCTGATGAGTGCACAAACCCCTATATATATGGCCCTGAGTATATCAAGTAC 435
 QY 103 PheAsnAspLysThrIleAspGluGluLeuGluAspLysArgValThrTTPIleVal 122
 Db 436 TTCATGATAAAACCATTTGATGAGAACTAGAACGGGACAAGAGGGTCACTTGGATTGTG 495
 QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
 Db 496 GAGTTCTTTGCCAATTTGGTCTAATGACTGCTCAATCATTTTGCCTCTATCTATGCTGACCTC 555
 QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
 Db 556 TCCCTTAATATACAACTGTACAGGCTAAATTTGGGAAGGTGGATGTGGACGCTATACT 615
 QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
 Db 616 GATGTTAGTACGGGTACAAAGTGAGCACATCACCTCCACCAAGCACTCCCTACCTG 675
 QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
 Db 676 ATCTCTTCCAAGTGGCAAGGAGCAATGCGGCGCCACAGATTGACAAAGAAGGACGG 735
 QY 203 AlaValSerTTPThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
 Db 736 GCTGCTCATGGACCTTCTCTGAGGAGATGTGATCCGAGAATTAACTTAAATGAGCTA 795
 QY 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal 242
 Db 796 TACCAGCGGCCCAAGAACTATCAAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTG 855
 QY 243 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysLysLys 258
 Db 856 GC-TCAACCCCCCACCACAGTGTGAGATGGGGGAAACAAGAAGGATAAA 902

RESULT 12
 AA242527
 ID AA242527 standard; cDNA; 774 BP.
 XX
 AC AA242527;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST isolated from a cDNA library SEQ ID NO:286.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification; ss.
 XX Homo sapiens.
 XX
 PN WO9953051-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 09-APR-1999; 99WO-IB00712.
 XX
 PR 09-APR-1998; 98US-0057719.
 PR 28-APR-1998; 98US-0069047.
 XX

Db 161 TGGCTCCGCCAACCTTACTACCTTCTGTGCGCCCTGCTCTCTGCTGCCCTTCCCTACTCG 220
Qy 39 aArgLysLeuProLeuCysHisGlyLeuProThrGln-ArgGluAspGlyAsn-Pr 58
Db 221 TGAGGAACATGCGCGCTGTGCGACGCTGTGCGACCCCAACCGAAGACGGTAACCC 280
Qy 58 oCysAspPheAspTrpArgGlu-ValGluIleLeuMetPheLeuSerAlaIleValMetM 78
Db 281 GTGTGACTTTGACTGGAGAAAGTGAGATCCTGTGATGTTTCTCAGTGCCCATGTGTGATGA 340
Qy 78 etLysAsn-ArgArgSer----- 83
Db 341 TGAAGACCCGACATCCATCAGTGGAGCAACATATAGCAACATTTTCATGTTTAAG 400
Qy 83 ----- 83
401 TAAAGTGGCCACACAACTTTCTTCCGCTTGGATATTCGATGGCGCTACTTTAC 460
Qy 84 -----MetPheLeuMetThrCysLys-----ProProLeuTyrMetGly 96
Db 461 ATCACACTGCTATAGTTTCTCATGACGTGCAACCCCTCCCTCCCA-TATATGGC 519
Qy 97 ProGluTyrIleLysTyrPheAsn-AspLysThrIleAspGluGluLeuGluArgAspLys 116
Db 520 CCTGAGTATATCACTACTCAATGATATAACCTTTGATGAGCACTAGAACGGGACAA 579
Qy 116 sArgValThrTrp-IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPhe 135
Db 580 CAGGTCATTTGGAATTTGGAGTCTTTGGCAATTTGGTCTAATGACTGGCAATCATTT 639
Qy 136 AlaProIleTyrAlaAspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLys 155
Db 640 GCCCTATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGCTAAATTTTGGGAG 699
Qy 156 ValAspValGlyArgTyrThrAspValSerThrArgTyrLysValSerThrSerProLeu 175
Db 700 GTGGATGTTGGACGCTACTATGATGTAGTACGGCGTACAAGTGAGCACATCACCCCTC 759
Qy 176 ThrLys-GlnLeuProThrLeu-IleLeuPheGlnGlyGlyLysGluAlaMet---ArgA 194
Db 760 ACCAAGGCAACTCCCTACCTGGATCTCTTCCAAAGTGGGCAAGGAGGCAATCGGGC 819
Qy 194 rgProGlnIle---AspLysLysGlyArg-AlaVal-SerTrpThrPhe----- 208
Db 820 GGCCACAGATTTGCACAGGAAGGAGCGGGCTGTTTTCATGGGACCTTCTCTGAGGTAC 879
Qy 209 -----SerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnAr 225
Db 880 CTGAAGGAAGGCGGAGGAGATGTATCCGAGAAATTAATTAATGAGCTATACCGCG 939
Qy 225 gAlaLysLysLeuSerLysAlaGlyAspAsnIle-ProGluGluGlnProValAlaSerT 245
Db 940 GGCAAGAAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCGCTGTGGCTTCAA 999
Qy 245 hrProThrValSerAsp-GlyGluAsnLysLysAspLys 258
Db 1000 CCCCCACAGTGTGATGGGGGAAACAAAGAGGATAAA 1041

RESULT 15

AAAA1964 standard; cDNA; 439 BP.

XX AC AAA41964;

XX XX 21-AUG-2000 (first entry)

XX DE Human secreted expressed sequence tag SEQ ID NO:704.

XX KW Human; mouse; xenopus; rat; secreted expressed sequence tag; sBST;

KW KW expressed sequence tag; EST; probe; chemotactic; proliferative;

KW KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

KW KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;

KW KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;

KW antilucer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX WO200021990-A1.
PN
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24205.
XX
PR 15-OCT-1998; 98US-0104435.
XX
PA (GENY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M;
XX
DR WPI; 2000-317937/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
XX expressed sequence tags (SESTs), useful for treating various disorders
XX such as autoimmune, infectious, and central nervous system disorders -
XX
XX Claim 1; Page 315; 618pp; English.

CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC cytokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytosatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antilucer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC in the exemplification of the present invention.

XX SQ Sequence 439 BP; 116 A; 111 C; 98 G; 114 T; 0 other;

Alignment Scores:

Pred. No.:	5,05e-66	Length:	439
Score:	622.00	Matches:	115
Percent Similarity:	99.15%	Conservative:	1
Best Local Similarity:	98.29%	Mismatches:	1
Query Match:	45.77%	Indels:	0
DB:	21	Gaps:	0

US-09-954-846-2 (1-258) x AAA41964 (1-439)

Qy 84 MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyrPhe 103

Db 83 GTGTCTCTGATGACGTGCAACCCCTATATATGCGGCTGTATATCAAGTACTTC 142

```
QY 104 AsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleValGlu 123
Db 143 RATGATAAAACCATTTGATGAGGAACCTAGAACGGGACAGAGGCTCACTTGGATTGGAG 202
QY 124 PhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSer 143
Db 203 TTCCTTGGCAATTGGTCTAATGACTGCCCAATCATTTGCCCTATCTATGCTGACCTCTCC 262
QY 144 LeuLysTyrAsnCysThrClyLeuAsnPheGlyLysValaspValGlyArgTyrThrAsp 163
Db 263 CTTAAATACAACTGTACAGGGCTAAATTTGGGAAGGTGGATGTTGGACGCTATACTGAT 322
QY 164 ValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrIle 183
Db 323 GTTAGTAGCGGTACAAAGTGAGCACAATCCCTCACCAGCACTCCCTACCCCTGATC 382
QY 184 LeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
Db 383 CTGTTCCAAGGTGGCAAGGAGCAATGCGGCGGCCACACAGATTGACAAATAAA 433
```

Search completed: July 12, 2003, 11:13:16
--b time : 191 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 12, 2003, 11:06:22 ; Search time 135 Seconds
(without alignments)
3017.635 Million cell updates/sec

Title: US-09-954-846-2

Perfect score: 1359

Sequence: 1 MAVLAPLIALYVPRLSRW.....EQPVASTPTVSDGENKKDK 258

~oring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet p2n.model -DEV=xlh
-O=/cn2.1/USPTO.spool/US0954846/runat_09072003_102146_12185/app_query.fasta_1.455
-DB=PublishedApplications_NA -QFWT=fastap -SUFFIX=rnpb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-NAXLEN=2000000000 -USER=US0954846.0CGN_1_1_125_0runat_09072003_102146_12185
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tabase : PublishedApplications_NA:

1: /cn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:
2: /cn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
6: /cn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:
7: /cn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1359	100.0	1555	10	US-09-954-846-4 Sequence 4, Appli
2	1359	100.0	1598	10	US-09-925-300-751 Sequence 751, App
3	1326	97.6	1620	9	US-09-905-291A-206 Sequence 206, App
4	1326	97.6	1620	9	US-09-902-853-206 Sequence 206, App

5	1326	97.6	1620	9	US-09-907-824-206	Sequence 206, App
6	1326	97.6	1620	9	US-09-907-841-206	Sequence 206, App
7	1326	97.6	1620	9	US-09-904-011-206	Sequence 206, App
8	1326	97.6	1620	9	US-10-174-590-31	Sequence 31, Appl
9	1326	97.6	1620	9	US-10-176-758-31	Sequence 31, Appl
10	1326	97.6	1620	9	US-10-175-737-31	Sequence 31, Appl
11	1326	97.6	1620	9	US-09-906-742-206	Sequence 206, App
12	1326	97.6	1620	9	US-10-173-706-31	Sequence 31, Appl
13	1326	97.6	1620	9	US-10-175-738-31	Sequence 31, Appl
14	1326	97.6	1620	9	US-10-175-752-31	Sequence 31, Appl
15	1326	97.6	1620	9	US-10-176-482-31	Sequence 31, Appl
16	1326	97.6	1620	9	US-10-176-757-31	Sequence 31, Appl
17	1326	97.6	1620	9	US-10-176-913-31	Sequence 31, Appl
18	1326	97.6	1620	9	US-10-180-552-31	Sequence 31, Appl
19	1326	97.6	1620	9	US-10-180-557-31	Sequence 31, Appl
20	1326	97.6	1620	9	US-09-906-838-206	Sequence 206, App
21	1326	97.6	1620	9	US-09-907-613-206	Sequence 206, App
22	1326	97.6	1620	9	US-09-907-942-206	Sequence 206, App
23	1326	97.6	1620	9	US-10-173-700-31	Sequence 31, Appl
24	1326	97.6	1620	9	US-10-174-572-31	Sequence 31, Appl
25	1326	97.6	1620	9	US-10-174-579-31	Sequence 31, Appl
26	1326	97.6	1620	9	US-10-174-582-31	Sequence 31, Appl
27	1326	97.6	1620	9	US-10-174-588-31	Sequence 31, Appl
28	1326	97.6	1620	9	US-10-175-739-31	Sequence 31, Appl
29	1326	97.6	1620	9	US-10-175-740-31	Sequence 31, Appl
30	1326	97.6	1620	9	US-10-175-743-31	Sequence 31, Appl
31	1326	97.6	1620	9	US-10-176-488-31	Sequence 31, Appl
32	1326	97.6	1620	9	US-10-176-492-31	Sequence 31, Appl
33	1326	97.6	1620	9	US-10-176-747-31	Sequence 31, Appl
34	1326	97.6	1620	9	US-10-176-750-31	Sequence 31, Appl
35	1326	97.6	1620	9	US-10-176-985-31	Sequence 31, Appl
36	1326	97.6	1620	9	US-10-176-987-31	Sequence 31, Appl
37	1326	97.6	1620	9	US-10-176-991-31	Sequence 31, Appl
38	1326	97.6	1620	9	US-10-176-992-31	Sequence 31, Appl
39	1326	97.6	1620	9	US-10-176-993-31	Sequence 31, Appl
40	1326	97.6	1620	9	US-10-184-658-31	Sequence 31, Appl
41	1326	97.6	1620	9	US-10-173-695-31	Sequence 31, Appl
42	1326	97.6	1620	9	US-10-173-697-31	Sequence 31, Appl
43	1326	97.6	1620	9	US-10-173-705-31	Sequence 31, Appl
44	1326	97.6	1620	9	US-10-174-576-31	Sequence 31, Appl
45	1326	97.6	1620	9	US-10-174-585-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-954-846-4
; Sequence 4, Application US/09954846
; Patent No. US20020102654A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Patterson, Chandra
; Baughn, Mariah R.
; TITLE OF INVENTION: THIOREDOXIN PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/954,846
; FILING DATE: 17-Sep-2001
; CLASSIFICATION: <Unknown>

QY 141 AspleuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValaspValGlyArg 160
Db 463 GACCTCCCTTAAATACAACTGACAGGGCTAATTTTGGAGAGGTGATGTTGGAGCC 522
QY 161 TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro 180
Db 523 TATACTGATGTTAGTACGGGTACAAAGTGAGACATCACCCCTCACCAAGCAACTCCCT 582
QY 181 ThrLeuIleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
Db 583 ACCCTGATCTGTCTCAAGGTGGCAAGGAGCAATGCGGGCCGACAGATTGACAAGAAA 642
QY 201 GlyArgAlaValSerTyrThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn 220
Db 643 GGAGGGCTGTCTCATGAGCTTCTGTGAGGAGATGTGATCCAGAAATTAATTAAT 702
QY 221 GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGln 240
Db 703 GAGCTATACCAGGGGGGCAAGAACTATCAAAAGGCTGGAGACAATATCCCTGAGGAGCAG 762
QY 241 ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
Db 763 CCTGTGGCTTCAACCCGCCACACAGTGTCTGAGATGGGAAAAACAAGAGGATAAA 816

RESULT 3

US-09-905-291A-206
; Sequence 206, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 206
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (973)
; OTHER INFORMATION: a, t, c or g
; NAME/KEY: modified_base
; LOCATION: (977)
; OTHER INFORMATION: a, t, c or g
; NAME/KEY: modified_base
; LOCATION: (996)
; OTHER INFORMATION: a, t, c or g
; NAME/KEY: modified_base
; LOCATION: (1003)
; OTHER INFORMATION: a, t, c or g
US-09-905-291A-206

Alignment Scores:

Pred. No.: 1.84e-175 Length: 1620
Score: 1326.00 Matches: 257
Percent Similarity: 87.16% Conservative: 1
Best Local Similarity: 86.82% Mismatches: 0
Query Match: 97.57% Indels: 38
DB: 9 Gaps: 1

US-09-954-846-2 (1-258) x US-09-905-291A-206 (1-1620)

QY 1 MetalValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTyr 20
Db 3 ATGGGGTCTTGGCACCTCTAATTGCTCGGTATTTCGGTGGCGGACTTTTCCAGATGG 62
QY 21 LeuAlaGlnProTyrTyrLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
Db 63 CTCGCCCAACCTTACTACTCTCTGCGCCCTCTCTGCTGCTCTCTCTCTCTCTCTCT 122
QY 41 LysLeuProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysasp 60
Db 123 AAATGCGCGCGCTCTGCCACGGTCTGCCACCCCAACGGAAGAGGTTACCGGTGTGAC 182
QY 61 PheAspTyrArgGluValIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db 183 TTTGACTGGAGAGAAGTGGAGATCTGATGTTTCTCAGTGCCATTGTGATGATGAAGAAC 242
QY 81 ArgArgSer-----83
Db 243 CGCAGATCCATCATCTGTGGAGCAACATATAGGCAACATTTTTCATGTTAGTAAAGTGCC 302
QY 83 -----83
Db 303 AACACAATCTTTTCTCCGCTTGGATATTCGATGGGCGCTACTTTTACATCACACTCTGC 362

Db 123 AAACGGCCGCGCTGCGCCAGGTGCGCCACCGGAAAGCGGTAACCCGCTGTGAC 182
Oy 61 PheAspTTPArqGluValGluLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db 183 TTTGACTGGAGAGAGTGGAGATCTCTGATGTTTCTCAGTGCCATTGTGATGATGAGAAC 242
Oy 81 ArgArgSer----- 83
Db 243 CGCAGATCCATCAGTGTGGAGCAACATAGGCAACATTTTCATGTTTAGTAAAGTGGCC 302
Oy 83 ----- 83
Db 303 AACACAAFTCTTTTCTCCGCTTGGATATTCGCATGGGCTACTTTACATCATCCTCTGC 362
Oy 84 ---MetPheLeuMetThrCysLysProPheLeuTyrMetGlyProGluTyrIleLysTyr 102
363 ATAGTGTCTCGATGACGTGCAACCCCTATATATGGGCCCTGAGTATATCAAGTAC 422
Oy 103 PheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleVal 122
Db 423 TTCAATGATAAACCATTTGATGAGGAACCTAGAACGGGACAGAGGCTCACTTGGATTGG 482
Oy 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrIleAspLeu 142
Db 483 GAGTTCTTTGGCAATTTGCTTAATGACTGCCAATCATTTGCCCTCTATCTATGCTGACCTC 542
Oy 143 SerLeuLysTyrAsnCysThrClyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
Db 543 TCCCTTAATAACAAGTACAGGGGCTAAATTTGGGAAGGTGGATTTGGAGCGCTATACT 602
Oy 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db 603 GATGTAGTACGGGTACAAGTGAGCACATCACCCCTCACCAAGCACTCCTACCTCG 662
Oy 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db 663 ATCCTGTTCAGGTGGCAAGGAGGCAATGCGCGCCACACAGATTGACAAGAAAGACGG 722
Oy 203 AlaValSerThrPheSerGluAlaValIleArgGluPheAsnLeuAsnGluLeu 222
Db 723 GCTGTCTCATGACCTCTCTGAGGAGAAATGTATCCGAGAATTTAACTTAAATGAGCTA 782
Oy 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal 242
783 TACCAGCGGGCCAGAACTATCAAGGCTGGAGACATATCCCTGAGGAGGAGCCTGTG 842
Oy 243 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
Db 843 GCTTCAACCCCGCCACACAGTGTGATGGGAAACAAAGAGATATAA 890

RESULT 7

US-09-904-011-206
; Sequence 206, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrata, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.

; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 206
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 973, 977, 996, 1003
; OTHER INFORMATION: unknown base
US-09-904-011-206
Alignment Scores:
Pred. No.: 1 84e-175 Length: 1620
Score: 1326.00 Matches: 257
Percent Similarity: 87.16% Conservative: 1
Best Local Similarity: 86.82% Mismatches: 0
Query Match: 97.57% Indels: 38
DB: 9 Gaps: 1
US-09-954-846-2 (1-258) x US-09-904-011-206 (1-1620)

Oy 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp 20
Db 3 ATGGCGGTCTTGGCACCTCTAATGCTCTCGTGATTCTCGTGGCGGCGGACTTTCACGATGG 62

```

; TITLE OF INVENTION:  ACIDS ENCODING THE SAME
; FILE REFERENCE:  P3430RIC42
; CURRENT APPLICATION NUMBER:  US/10/174,590
; CURRENT FILING DATE:  2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS:  612
; SEQ ID NO 31
; LENGTH:  1620
; TYPE:  DNA
; ORGANISM:  Homo Sapien
; FEATURE:
; NAME/KEY:  unsure
; LOCATION:  973, 977, 996, 1003
; OTHER INFORMATION:  unknown base
US-10-174-590-31

Alignment Scores:
Pred. No.:      1,84e-175      Length:      1620
Score:          1326.00      Matches:      257
Percent Similarity:  87.16%      Conservative:  1
Best Local Similarity:  86.82%      Mismatches:   0
Query Match:       97.57%      Indels:       38
DB:               9          Gaps:         1

US-09-954-846-2 (1-258) x US-10-174-590-31 (1-1620)

Qy      1  MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrrp 20
Db      3  ATGGCGGCTTGGCACCTCTAATTCCTCTCGTGATTCGGTCCGCGACTTCACCATGG 62
Qy     21  LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
Db     63  CTCGCCCAACCTTACTACCTTCTGTGCGCCCTGCTCTCTGCTGCCTTCTACTCTGTGAGG 122
Qy     41  LysLeuProProLeuCyHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
Db    123  AAACCTGCCGCGCTCTGCCACGGTCTGCCACCCCAACGCGAAGACGGTAACCCGTGTGAC 182
Qy     61  PheAspTrrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db    183  TTTGACTCGAGAAAGTGGAGATCTCGATGTTTCAGTGCCCATTTGTGATGATGAAGAAC 242
Qy     81  ArgArgSer----- 83
Db    243  CGCAGATCCACTCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGCC 302
Qy     83  ----- 83
Db    303  AACACAATTTCTTTCTCCGCTTGGATATTCGATGGGCTACTTTACATCACACTCTGC 362
Qy     84  --MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTrrIleLysTyr 102
Db    363  ATAGTGTTCCGTGATGACGTGCAGAACCCCCCTATATATGGCCCTGAGTATATCAAGTAC 422
Qy    103  PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValTrrIrrIleVal 122
Db    423  TTCAATGATAAAACCATTTGATGAGGAACATAGAACGGGACAGAGGGCTCACTTGGATTGTG 482
Qy    123  GluPhePheAlaAsnTrrpSerAsnAspCysGlnSerPheAlaProIleTrrAlaAspLeu 142
Db    483  GAGTTCTTTGGCAATTGGCTTAATGACTGCCAATCATTTGCCCTTATCTATGCTGACCTC 542
Qy    143  SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTrrThr 162
Db    543  TCCCTTAAATCAACTGTACAGGGCTAAATTTTGGAAAGGTGGATGTTGGAGCGCTATACT 602
Qy    163  AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db    603  GATGTTAGTACCGGTACAAAGTGGAGCAATCACCCCTCACCAAGCAACTCCCTACCCCTG 662
Qy    183  IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db    663  ATCTGTTCACAAAGTGGCAAGAGGCATATCGCGCGCCACACATTCACAGAAAGCAGCG 722

```


Qy	83	-----	83	Alignment Scores:	1.84e-175	Length:	1620
				Pred. No.:	1326.00	Matches:	257
				Score:	87.16%	Conservative:	1
				Percent Similarity:	86.82%	Mismatches:	0
				Best Local Similarity:	97.57%	Indels:	38
				Query Match:	9	Gaps:	1
				DB:			
Qy	83	-----	83	US-09-954-846-2 (1-258) x US-10-173-706-31 (1-1620)			
Db	303	AACACAAATCTTTCTCCGCTTGGATATCGCATGGCGCTACTTTTACATCACACTCTGC	362	Qy 1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTTP			
Qy	84	--MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr	102	Db 3 ATGCGGGTCTTGGCACCTCTAAATTCGCTCTGCTGATTCGGTCCGCGACATTCACGATGG			
Qy	363	ATAGTGTCTCTGATGAGGTGCAAAACCCCTATATATGGCCCTGAGTATATCAAGTAC	422	Qy 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg			
Qy	103	PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrIleVal	122	Db 63 CTCGCCAACCTTACTACCTTCTCTCGGCCCTGCTCTCTGCTTCTCTACTCTGCTGAG			
Db	423	TTCAATGATAAAACCATTTGATGAGGAACTAGACGGGACAAAGGCTCACTTGGATTGG	482	Qy 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp			
Qy	123	GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu	142	Db 123 AAATCGCGCGCTCTGCCACGGTCTGCCACCCACCAACGCGAAGACGGTAACCGTGTGAC			
Db	483	GAGTTCTTCCCAATTTGGTCTAATGATGCGCAATCATTTGCCCTATCTATGCTGACCTC	542	Qy 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn			
Qy	143	SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr	162	Db 183 TTTCACTGGAGAAAGTGGAGATCCTGATGTTTCTCAGTCGCCATTTGGATGATGAAGAAC			
Db	543	TCCCTTAATACAACTGTACAGGGCTAAATTTGGGAAGGTGGATGTTGGACGCTATACT	602	Qy 81 ArgArgSer-----			
Qy	163	AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu	182	Db 243 CGCAGATCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGCC			
Db	603	GATGTTAGTACGGGTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCCCTG	662	Qy 83 -----			
Qy	183	IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg	202	Db 303 AACCAAAATCTTTTCTCCGCTTGGATATTCGCATATTCGCATGGGCGCTACTTTATCATCACACTCTGC			
Db	663	ATCCTGTTCAGGTGGCAAGGAGGAATTCGCGGCCACAGATTCACAAGAAAGACGG	722	Qy 83 -----			
Qy	203	AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu	222	Db 302			
Db	723	GCTGCTCATGGACCTCTCTGAGGAGATGTCATCCGAGAATTTAACTTAATGAGCTA	782	Qy 83 -----			
Qy	223	TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal	242	Db 302			


```

; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 31
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 973, 977, 996, 1003
; OTHER INFORMATION: unknown base
; US-10-175-752-31

Alignment Scores:
Pred. No.: 1,84e-175      Length: 1620
Score: 1326.00           Matches: 257
Percent Similarity: 87.16% Conservative: 1
Best Local Similarity: 86.82% Mismatches: 0
Query Match: 97.57%      Indels: 38
DB: 9                    Gaps: 1

US-09-954-846-2 (1-258) x US-10-175-752-31 (1-1620)

Qy 1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTrp 20
Db 3 ATGGCGGTCTTGGCACTTAATTCGTCTCGTGTATTCGTCGCGGCACTTTCACGATGG 62
Qy 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuSerAlaAlaPheLeuValArg 40
Db 63 CTGCGCCAACTTACCTTCTGTCGCGGCTCTCTGTCGCGCTTCTACTCTGTGAGG 122
Qy 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
Db 123 AAACGCGCGCTCTGCGACGCTGTCGCCACCCACGCGAGACGGTAACCCGCTGAC 182
Qy 61 PheAspTrpArgGluValGluLeuLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db 183 TTTGACTGGAGAGAAAGTGGAGATCCTGATGTTTCTCAGTGCCATGTGATGAAGAAC 242
Qy 81 ArgArgSer----- 83
Db 243 CGCAGATPCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGCC 302
Qy 83 ----- 83
Db 303 AACACAATCTTTTCTTCGCTGGATATTCGCATCGGCTACTTTACATCACTCTGC 362
Qy 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
Db 363 ATAGTGTCTGTATGACGTGCAAAACCCCTATATATATGCGCCCTGAGTATCAAGTAC 422
Qy 103 PheAsnAspLysThrIleAspGluLeuLeuGluArgAspLysArgValThrIleVal 122
Db 423 TTAAGTATATAAACCATTCATGAGGAACATAGAACGGGCAAGAGGGTCTACTTGGATTGTG 482
Qy 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db 483 GAGTCTTTGCCAATTTGGTCTATGACTGCCAATCATTTGCCCTTATCTATCTGACCTC 542
Qy 143 SerLeuLysTyrAsnCysThrGlyLeuLeuAsnPheGlyLysValAspValGlyArgTyrThr 162

```

[illegible]

```
QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
Db 123 AAACAGCCGCCCTCTGCCACCGCTGTGCCACCAACGCAAGAGCGGTAAACCGTGTGAC 182
QY 61 PheAspTTPArgGluValGluLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db 183 TTTGACTGGAGAGAGTGGAGATCCTGATGTTCTCAGTGCCCATCTGTGATGATGAAGAAC 242
QY 81 ArgArgSer----- 83
Db 243 CGCAGATCCATCACTGTGGAGCAACATATAGGCACACATTTTCATGTTTAGTAAAGTGCC 302
QY 83 ----- 83
Db 303 AACACAAATCTTTCTTCTCGCTTGGATATTCGCATGGGCTACTTTACATCACACTCTGC 362
QY 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
Db 363 ATAGTGTTCTCTCATGACGTGCAACCCCTTATATATGGCCCTCAGTATATCAAGTAC 422
QY 103 PheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleVal 122
Db 423 TTCAATGATAAAACCATTCATGAGGAACTAGAACGGGACAAAGAGGCTCACTGGATTGTG 482
QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db 483 GAGTTCTTTGGCAATTTGGTCTAATGACTGCCAATCTTTGCCCTATCTATGCTGACCTC 542
QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
Db 543 TCCTTAAATACAACTGTACAGGGCTAAATTTGGGAAGGTGGATGTTGGACCTATACT 602
QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db 603 GATGTTAGTACCGGTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCCCTG 662
QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db 663 ATCCCTGTTTCCAAAGGTGGCAAGGAGGCAATCGCGCGGCCACAGATTGCACAAAGAGACGG 722
QY 203 AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
Db 723 GCTGTCTCATGGACCTTCTCTGAGGAGATGTGATCCGAGATTTAACTTAAATGAGCTA 782
QY 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal 242
Db 783 TACCAGCGGGCCCAAGAACTATCAAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTG 842
QY 243 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
Db 843 GUTTCACCCCCACCACAGTGTGATGGGGGAAACAAAGAGGATAAA 890
```

Search completed: July 12, 2003, 12:06:07
Job time : 142 secs

Db 26 CGGTTACGGCGGAAAGATGGCGGTCTTGACACTCTAATGTCTCGTGTATTCGTGC 85
QY 102 CGCGACTTTACAGATGGCTGCGCCAACTTACCTTACCTTCTGTGCGGCCCTCTCTCTGCTG 161
Db 86 CGCGACTTTACAGATGGCTGCGCCAACTTACCTTACCTTCTGTGCGGCCCTCTCTCTGCTG 145
QY 162 CCTTCTACTCTGTGAGAACTGCGCGCTCTGCGACGGTCTGCGCCCAACCAAGCGAAG 221
Db 146 CCTTCTACTCTGTGAGAACTGCGCGCTCTGCGACGGTCTGCGCCCAACCAAGCGAAG 205
QY 222 ACGGTAACCGCTGTGACTTGTGAGAGAACTGAGAGTCTGATGTTTCTCAGTGCCA 281
Db 206 ACGGTAACCGCTGTGACTTGTGAGAGAACTGAGAGTCTGATGTTTCTCAGTGCCA 265
QY 282 TTGTGATGATGAAGAACCGCAGATCCATGTTCTGTGAGTGCGCAAAACCCCTCTATATA 341
Db 266 TTGTGATGATGAAGAACCGCAGATCCATGTTCTGTGAGTGCGCAAAACCCCTCTATATA 325
QY 342 TGGGCCCTGAGTATATCAAGTACTTCAATGATAAACCAATGTGAGGAACTAGAAGCGG 401
Db 326 TGGGCCCTGAGTATATCAAGTACTTCAATGATAAACCAATGTGAGGAACTAGAAGCGG 385
QY 402 ACAAGAGGGTCACTTGGATTGTGGAGTCTTTCGCAATTGGTCTAATGACTGCCAATCAT 461
Db 386 ACAAGAGGGTCACTTGGATTGTGGAGTCTTTCGCAATTGGTCTAATGACTGCCAATCAT 445
QY 462 TTGCCCTTATCTATGCTGACCTTCCCTTAAATACAACTGACAGGCTAAATTTTGGGA 521
Db 446 TTGCCCTTATCTATGCTGACCTTCCCTTAAATACAACTGACAGGCTAAATTTTGGGA 505
QY 522 AGGTGATGTTGACGCTATATGATGTAGTAGGGGTACAAAGTGAGCACATCACCC 581
Db 506 AGGTGATGTTGACGCTATATGATGTAGTAGGGGTACAAAGTGAGCACATCACCC 565
QY 582 TCACCAAGCAACTCCCTACCTCATCTGCTTCAAGGTGCAAGGAGCAATGCGCGGC 641
Db 566 TCACCAAGCAACTCCCTACCTCATCTGCTTCAAGGTGCAAGGAGCAATGCGCGGC 625
QY 642 CACAGATTGACAAAGAGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTATCC 701
Db 626 CACAGATTGACAAAGAGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTATCC 685
QY 702 GAGAAATTAATTAATGAGCTATACAGCGGCGCAAGAACTATCAAGGCTGGAGACA 761
Db 686 GAGAAATTAATTAATGAGCTATACAGCGGCGCAAGAACTATCAAGGCTGGAGACA 745
QY 762 ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACAGTGTCAAGTGGGGAACA 821
Db 746 ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACAGTGTCAAGTGGGGAACA 805
QY 822 AGAAGGATAAATAAGATCCCTCACTTTGGCAGTGTCTCTCTCTGTCCTCAATTCAGGCTCT 881
Db 806 AGAAGGATAAATAAGATCCCTCACTTTGGCAGTGTCTCTCTCTGTCCTCAATTCAGGCTCT 865
QY 882 TTCCATAACACAGGCTGAGGCTGACGCTTTTATTTATGTTTTCCTTTGCGCTGTGAC 941
Db 866 TTCCATAACACAGGCTGAGGCTGACGCTTTTATTTATGTTTTCCTTTGCGCTGTGAC 925
QY 942 TGGGTGGGCGAGCATGACGCTTCTGATTTTAAAGAGGCATCTAGGGAAATGTACAGGCAC 1001
Db 926 TGGGTGGGCGAGCATGACGCTTCTGATTTTAAAGAGGCATCTAGGGAAATGTACAGGCAC 985
QY 1002 CTACAGGAAGGCTGCGCATGCTGTGGCAACTGTTTTCACCTGGAGCAAGAAAGAGATCTCA 1061
Db 986 CTACAGGAAGGCTGCGCATGCTGTGGCAACTGTTTTCACCTGGAGCAAGAAAGAGATCTCA 1045
QY 1062 TAGGACGGAGGGGAAATGGTTTCCCTTCCAAAGCTTGGGTGAGTGTGTTAACTGCTTATCA 1121
Db 1046 TAGGACGGAGGGGAAATGGTTTCCCTTCCAAAGCTTGGGTGAGTGTGTTAACTGCTTATCA 1105
QY 1122 GCTATTTCAGACATCTCCATGGTTTCTCCATGAACACTCTGTGGTTTCATCATCTCTCTTA 1181

Db 1106 GCTATTTCAGACATCTCCATGGTTTCTCCATGAACACTCTGTGGTTTTCATCATCTCTCTTA 1165
QY 1182 GTTGACCTTCACAGCTTGGTTAGACCTTAGATTTAAACCTTAAGGTAAGATGCTGGGGATA 1241
Db 1166 GTTGACCTTCACAGCTTGGTTAGACCTTAGATTTAAACCTTAAGGTAAGATGCTGGGGATA 1225
QY 1242 GAACGCTTAAGAATTTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCTTGGCTTCGTTTA 1301
Db 1226 GAACGCTTAAGAATTTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCTTGGCTTCGTTTA 1285
QY 1302 TGTGCTTCATTAAGAATTAAGCCTAATTTGTGCTAGTCTTAAAGGAAACCTTTAAC 1361
Db 1286 TGTGCTTCATTAAGAATTAAGCCTAATTTGTGCTAGTCTTAAAGGAAACCTTTAAC 1345
QY 1362 CACAAAGTTTTATCATTTGAAGCAATATTGAACAAACCCCTATTCTTGGGGATTGAGA 1421
Db 1346 CACAAAGTTTTATCATTTGAAGCAATATTGAACAAACCCCTATTCTTGGGGATTGAGA 1405
QY 1422 AGGGTCAATAGAGGCTTTGAGACTTCTTGTGTGTAGGACTTGGAGGAGAAATCCCC 1481
Db 1406 AGGGTCAATAGAGGCTTTGAGACTTCTTGTGTGTAGGACTTGGAGGAGAAATCCCC 1465
QY 1482 TGGACTTTCATAACCTCTGACATACTCCCAACACCCAGTTGATGGCTTTCCGTAATAA 1541
Db 1466 TGGACTTTCATAACCTCTGACATACTCCCAACACCCAGTTGATGGCTTTCCGTAATAA 1525
QY 1542 AAAGATTGGGATT 1554
Db 1526 AAAGATTGGGATT 1538

RESULT 3

US-09-852-797-19
; Sequence 19, Application US/09852797
; Patent No. US20020172994A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P2
; CURRENT APPLICATION NUMBER: US/09/852,797
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 19:
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE

Qy	32	CGAGCAGTGGCCGTTACCGCGCAAAAGATGGCGGCTTTGGCACCTCTAATTTGCTCTCGTG	91
Db	3	CGTGGCAATTTCCGACGAGCAAAAGATGGCGGCTTTGGCACCTCTAATTTGCTCTCGTG	92
Qy	92	TATTTCGGTGGCGGACTTTCACGATGGGTCGCGCCAAACCTTACTACTCTCTGTGGCCCGTG	151
Db	63	TATTTCGGTGGCGGACTTTCACGATGGGTCGCGCCAAACCTTACTACTCTCTGTGGCCCGTG	122
Qy	152	CTCTCTGCTGCTTTCCTACTCGTGTAGGAAATGCGCGCGCTCTGTGCCACGGTCTGCCACAC	211
Db	123	CTCTCTGCTGCTTTCCTACTCGTGTAGGAAATGCGCGCGCTCTGTGCCACGGTCTGCCACAC	182
Qy	212	CAACGCGAAGACGCTAACCCCTGTGTACCTTGTACTGTGGAGAGAGTGGAGATCCTGATGTTT	271
Db	183	CAACGCGAAGACGCTAACCCCTGTGTACCTTGTACTGTGGAGAGAGTGGAGATCCTGATGTTT	242
Qy	272	CTCAGTGCATTTGTGATGATGAAGAACCGCAGATCCA	308
Db	243	CTCAGTGCCATTTGTGATGATGAAGAACCGCAGATCCA	302
Qy	309	-----	308
Db	303	AACATTTTTCATGTTAGTAAAGTGGCCAAACAATTTCTTCTCGCTTGGATATTCGC	362
Qy	309	-----TGTTCTCTGATCAGCTGCAACCCCGCCCTA	337
Db	363	ATGGGCGCTACTTTACATCACACTCTGCATAGTTCTCTGATCAGCTGCAACCCCGCCCTA	422
Qy	338	TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAAACCATTTGATGAGGAAGTAGAA	397
Db	423	TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAAACCATTTGATGAGGAAGTAGAA	482
Qy	398	CGGACAAGAGGGTCACCTTGGAATTTGGAGTTCTTTTGGCAATTTGGTCTAATGACTGCCAA	457
Db	483	CGGACAAGAGGGTCACCTTGGAATTTGGAGTTCTTTTGGCAATTTGGTCTAATGACTGCCAA	542
Qy	458	TCATTTGCCCTATCTATGCTGACCTCTCCCTTAAATACAAGTGTACAGGGCTAAATTTT	517
Db	543	TCATTTGCCCTATCTATGCTGACCTCTCCCTTAAATACAAGTGTACAGGGCTAAATTTT	602
Qy	518	GGGAAGTGGATGTTGGACGCTATACTGATGTTAGTACGCGGTACAAAAGTGAGCACATCA	577
Db	603	GGGAAGTGGATGTTGGACGCTATACTGATGTTAGTACGCGGTACAAAAGTGAGCACATCA	662
Qy	578	CCCCACCAAGCAACTCCCTACCCCTGATCTGTTTCCAAAGTGGCAAGGAGCAATGCGG	637
Db	663	CCCCACCAAGCAACTCCCTACCCCTGATCTGTTTCCAAAGTGGCAAGGAGCAATGCGG	722
Qy	638	CGGCCACAGATTGACAGAAAGGAGGGGCTCTCATGACCTTCTCTGAGGAGATGTG	697
Db	723	CGGCCACAGATTGACAGAAAGGAGGGGCTCTCATGACCTTCTCTGAGGAGATGTG	782
Qy	698	ATCCGAGAAATTTAACTTAAATGAGCTATACCAGCGGGGCCAGAAAACCTATCAAAAGCTGGA	757
Db	783	ATCCGAGAAATTTAACTTAAATGAGCTATACCAGCGGGGCCAGAAAACCTATCAAAAGCTGGA	842
Qy	758	GACAAATATCCCTGAGGAGCACCTGTGGCTTCAACCCCGCCACACAGTGTGATGGGAA	817
Db	843	GACAAATATCCCTGAGGAGCACCTGTGGTGTTCACCCCGCCACACAGTGTGATGGGAA	902
Qy	818	AACAAGAGGATAAATAAGATCCCTACCTTTGGCAGTGTCTCTCTGCTCAATTTCCACAG	877
Db	903	AACAAGAGGATAAATAAGATCCCTACCTTTGGCAGTGTCTCTCTGCTCAATTTCCACAG	962
Qy	878	CTCTTTTCCATAACCCAAAGCCTGAGGGTGCAGCCTTTTATTTATCTTTTCCCTTTGGGTG	937

```

RESULT 4
US-09-853-161-19
; Sequence 19, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P
; CURRENT APPLICATION NUMBER: US/09/853,161
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357

```

; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (871)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-853-161-19

Query Match 88.8%; Score 1381.6; DB 10; Length 1699;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1511; Conservative 1; Mismatches 11; Indels 114; Gaps 1;

QY 32 CGAGCAGTGGCGTTACGCCCGGAAAGATGGCGGCTTGGCCACCTCTAATTGCTCTCGTG 91
DB 3 CGTGGCGAATTGGGCACGAGCGAAAGATGGCGGCTTGGCCACCTCTAATTGCTCTCGTG 62
QY 92 TATTCGGTGGCGGACTTTACGATGGCTCGGCCAACCTTACTACCTTCTGTGGCGCCCTG 151
DB 63 TATTCGGTGGCGGACTTTACGATGGCTCGGCCAACCTTACTACCTTCTGTGGCGCCCTG 122
QY 152 CTCCTGTGCTTCTCTACTGTGAGAAACTGCCGCCCTCTGCCACGCTGTGCCACC 211
DB 123 CTCCTGTGCTTCTCTACTGTGAGAAACTGCCGCCCTCTGCCACGCTGTGCCACC 182
QY 212 CAACGGAGACGGTAACCCGTGTGACTTTGACTGGAGAGAGTGGAGATCTGTATGTTT 271
DB 183 CAACGGAGACGGTAACCCGTGTGACTTTGACTGGAGAGAGTGGAGATCTGTATGTTT 242
QY 272 CTCAGTGCCATTGTGATGATGAAGAACCGCAGATCCA----- 308
DB 243 CTCAGTGCCATTGTGATGATGAAGAACCGCAGATCCA----- 302
QY 309----- 308
DB 303 AACATTTTCATGTTAGTAAAGTGGCCAAACACAAATCTTTCTTCGCTTGGATATTCGC 362
QY 309-----TGTTCTCTGATGACGTGCAACCCCTTCA 337
DB 363 ATGGCCCTACTTTACATCACACTCTGCATAGTGTCTCTGATGACGTGCAACCCCTTCA 422
QY 338 TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACCTAGAA 397
DB 423 TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACCTAGAA 482
QY 398 CGGGACAAGAGGGTCACTTGGATTTGGGAGTTCTTTGGCCAAATTTGCTAATGACTGCCAA 457
DB 483 CGGGACAAGAGGGTCACTTGGATTTGGGAGTTCTTTGGCCAAATTTGCTAATGACTGCCAA 542
QY 458 TCATTTGGCCCTTATCTATGCTGACCTCTCCCTTAATACAACTGTACAGGGCTAAATTTT 517
DB 543 TCATTTGGCCCTTATCTATGCTGACCTCTCCCTTAATACAACTGTACAGGGCTAAATTTT 602
QY 518 GGGAGGTGGATGTTGGAGCTTATCTATGTTAGTACGGGTACAAAGTGGAGCATCA 577
DB 603 GGGAGGTGGATGTTGGAGCTTATCTATGTTAGTACGGGTACAAAGTGGAGCATCA 662
QY 578 CCCCTCACCAGCAACTCCCTACCTGATPCCTGTTTCCAAAGGTGGCAAGGAGCAATGCGG 637
DB 663 CCCCTCACCAGCAACTCCCTACCTGATCCTGTTTCCAAAGGTGGCAAGGAGCAATGCGG 722

RESULT 5

US-09-852-659A-19
; Sequence 19, Application US/09852659A
; Patent No. US2002007287A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P4

QY 638 CGGCCACAGATTGACAAGAAAGGACGGGCTGTCTCATGGACCTTCTCTGAGGAGAATGTG 697
DB 723 CGGCCACAGATTGACAAGAAAGGACGGGCTGTCTCATGGACCTTCTCTGAGGAGAATGTG 782
QY 698 ATCCGAGAAATTTAACTTAAATGAGCTATATACCGAGCGGCCAAGAACTATCAAGGCTGGA 757
DB 783 ATCCGAGAAATTTAACTTAAATGAGCTATATACCGAGCGGCCAAGAACTATCAAGGCTGGA 842
QY 758 GACAATATCCCTGAGGAGGAGCCTGTGGCTTCAACCCGCCACCCAGTGTCTAGATGGGAA 817
DB 843 GACAATATCCCTGAGGAGGAGCCTGTGGCTTCAACCCGCCACCCAGTGTCTAGATGGGAA 902
QY 818 AACAAGAAGGATAAATAAGATCTCACTTTGGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 877
DB 903 AACAAGAAGGATAAATAAGATCTCACTTTGGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 962
QY 878 CTCCTTTCCATAACCAACCAAGCCTGAGGCTGAGGCTTTTATTTATTTTCCCTTTGGCTG 937
DB 963 CTCCTTTCCATAACCAACCAAGCCTGAGGCTGAGGCTTTTATTTATTTTCCCTTTGGCTG 1022
QY 938 TGACTGGGTGGGCGAGCATCGAGCTTCTGATTTTAAAGAGGATCTTAGGGAATTTGTCAGG 997
DB 1023 TGACTGGGTGGGCGAGCATCGAGCTTCTGATTTTAAAGAGGATCTTAGGGAATTTGTCAGG 1082
QY 998 CACCCTACAGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1057
DB 1083 CACCCTACAGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1142
QY 1058 CTCATAGGCGAGGAGGAAATGTTTCCCTCAAGCTTGGGTCAGTGTGTAACTGCTT 1117
DB 1143 CTCATAGGCGAGGAGGAAATGTTTCCCTCAAGCTTGGGTCAGTGTGTAACTGCTT 1202
QY 1118 ATCAGCTATTCAGACATCTCCATGGTTCCTTCCATGAAACTCTGTGGTTCATCATTCCTT 1177
DB 1203 ATCAGCTATTCAGACATCTCCATGGTTCCTTCCATGAAACTCTGTGGTTCATCATTCCTT 1262
QY 1178 CTTAGTTGACCTGCACAGCTTGGTTAGACCTAGATTTAAACCTTAAGCTAAGATGCTGGG 1237
DB 1263 CTTAGTTGACCTGCACAGCTTGGTTAGACCTAGATTTAAACCTTAAGCTAAGATGCTGGG 1322
QY 1238 TATAGAAGCGTAAGAATTTTCCCCCAAGGACTCTTGTCTTCTTAAAGCCCTTCTGGCTTCG 1297
DB 1323 TATAGAAGCGTAAGAATTTTCCCCCAAGGACTCTTGTCTTCTTAAAGCCCTTCTGGCTTCG 1382
QY 1298 TTTATGCTCTTCATTTAAAGTATAAGCCTTAACCTTTGTCGCTAGTCTTAAAGAGAAACCTT 1357
DB 1383 TTTATGCTCTTCATTTAAAGTATAAGCCTTAACCTTTGTCGCTAGTCTTAAAGAGAAACCTT 1442
QY 1358 TAACCACAAGTTTTATCATTTGAAGACAATATTGAACACCCCTTATTTTGTGGGATTT 1417
DB 1443 TAACCACAAGTTTTATCATTTGAAGACAATATTGAACACCCCTTATTTTGTGGGATTT 1502
QY 1418 GAGAAGGGTGAATAGAGGCTTGAAGCTTTCCCTTTGTGTGTAGGACTTGGAGGAGAAAT 1477
DB 1503 GAGAAGGGTGAATAGAGGCTTGAAGCTTTCCCTTTGTGTGTAGGACTTGGAGGAGAAAT 1562
QY 1478 CCCTTGAGCTTTCACCTTAACCCCTCTGACATACTCCACACCCAGTGTGATGGCTTTCCGTA 1537
DB 1563 CCCTTGAGCTTTCACCTTAACCCCTCTGACATACTCCACACCCAGTGTGATGGCTTTCCGTA 1622
QY 1538 AAAAAAGATTGGGATT 1554
DB 1623 AAAAAAGATTGGGATT 1639

```

; CURRENT APPLICATION NUMBER: US/09/852, 659A
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (871)
; OTHER INFORMATION: n equals a.t.g. or c
US-09-852-659A-19

```

Query Match	88.8%;	Score 1381.6;	DB 10;	Length 1699;
Best Local Similarity	92.3%;	Pred. No. 0;		
Matches 1511;	Conservative 1;	Mismatches 11;	Indels 114;	Gaps 1
QY	32	CGAGCAGTGGCCGTTACGGCCGAAAGATGCGGGTCTTTGGCACCTCTAATGCTCTCGTG	91	
DB	3	CGTGCCGAATTGCGCACGAGCGAAAGATGCGGTCTTTGGCACCTCTAATTTGCTCTCGTG	62	
QY	92	TATTCGGTGGCGGACATTTCACGATGCGTCGCCCAACCTTACTACCTTCTGTGCGGCCCTG	151	
DB	63	TATTCGGTGGCGGACATTTCACGATGCGTCGCCCAACCTTACTACCTTCTGTGCGGCCCTG	122	
QY	152	CTCTCTGCTGCCTTCCCTACTCGTGAGGAAACTGCGCGCGCTCTGCGACGGTCTGCCAC	211	
	123	CTCTCTGCTGCCTTCCCTACTCGTGAGGAAACTGCGCGCGCTCTGCGACGGTCTGCCAC	182	
QY	212	CAACGGGAAGCGGTAAACCGTGTGACTTTGACTGGAGAGAGTGAGGAGATCCTGATGTTT	271	
DB	183	CAACGGGAAGCGGTAAACCGTGTGACTTTGACTGGAGAGAGTGAGGATCCTGATGTTT	242	
QY	272	CTCAGTGCCATTGTGATGATCAAGAACCGCAGATCCA-----	308	
DB	243	CTCAGTGCCATTGTGATGATGAAGAACCGCAGATCCATCCTGTTGGAGCAACATATAGGC	302	
QY	309	-----	308	
DB	303	AACATTTTCATGTTAGTAAAGTGGCCCAACACAATTTCTTTCTCCCGCTTGGATATTCG	362	
QY	309	-----TGTTCTGTGATGACGTGCAAAACCCCCCCTA	337	
DB	363	ATGGGCCTACTTTACATCACACTCTGCATAGTGTCTGTGACGTGCAAAACCCCCCCTA	422	
QY	338	TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACTAGAA	397	
DB	423	TATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACTAGAA	482	

Qy	603	TGATCTCTGTTCCAAAGGTGGCAAGGAGGCAATGCGGGGGCCACAGATTGACAGAAAGGAC	662
Db	661	TGATCTCTGTTCCAAAGGTGGCAAGGAGGCAATGCGGGGGCCACAGATTGACAGAAAGGAC	720
	663	GGGCTGCTCATGAGACCTTCTCTGAGGAGAAATGTATCCGAGAAATTTAACTTTAAATGAGC	722
Db	721	GGGCTGCTCATGAGACCTTCTCTGAGGAGAAATGTATCCGAGAAATTTAACTTTAAATGAGC	780
Qy	723	TATACCAAGCGGGCCAAAGAACTATCAAGAGCTGGAGACAATATCCCTTGAGGAGCAGCCTG	782
Db	781	TATACCAAGCGGGCCAAAGAACTATCAAGAGCTGGAGACAATATCCCTTGAGGAGCAGCCTG	840
Qy	783	TGGCTTCAACCCCCACACACAGTGTACATGGGGAACAAAGAGGATAAATAAGATCCTC	842
Db	841	TGGCTTCAACCCCCACACACAGTGTACATGGGGAACAAAGAGGATAAATAAGATCCTC	900
Qy	843	ACTTTGGCAGTGTCTTCTCTCTCTCAATGCCAGGCTCTTCCATAACCAACAGCCCTGAG	902
Db	901	ACTTTGGCAGTGTCTTCTCTCTCTCAATGCCAGGCTCTTCCATAACCAACAGCCCTGAG	960
Qy	903	GCTGCAGCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGACGATGCAAGCT	962
Db	961	GCTGCAGCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGACGATGCAAGCT	1020
Db	963	TCGTGATTTTAAAGAGGCACTAGGGAAATGTTCAGGCACCTTACAGGAAGGCTTGCCTATGC	1022
	1021	TCGTGATTTTAAAGAGGCACTAGGGAAATGTTCAGGCACCTTACAGGAAGGCTTGCCTATGC	1080
Qy	1023	TGTGGCCAACTGTTTCACTGGAGCAAGAAAGAGATCTCATAGGACGAGGGGGAATGGT	1082
Db	1081	TGTGGCCAACTGTTTCACTGGAGCAAGAAAGAGATCTCATAGGACGAGGGGGAATGGT	1140
Qy	1083	TTCCCTCCAAGCTTGGGTCAGTGTGTTAACTTGCTTATCAGCTATTACAGACATCTCCATGG	1142
Db	1141	TTCCCTCCAAGCTTGGGTCAGTGTGTTAACTTGCTTATCAGCTATTACAGACATCTCCATGG	1200
Qy	1143	TTTCTCCATGAACCTCTGTGGTTTCATCTTCCTTCTTAGTTGACCTGACAGCTTGGTT	1202
Db	1201	TTTCTCCATGAACCTCTGTGGTTTCATCTTCCTTCTTAGTTGACCTGACAGCTTGGTT	1260
Qy	1203	AGACCTAGATTTTAAACCTTAAGTAAAGATGCTGGGGTATAGAACCTTAAGAAATTTTCCCCC	1262
Db	1261	AGACCTAGATTTTAAACCTTAAGTAAAGATGCTGGGGTATAGAACCTTAAGAAATTTTCCCCC	1320
Qy	1263	AAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTATGGCTTCATTAAAGATATAA	1322
Db	1321	AAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTATGGCTTCATTAAAGATATAA	1380
Qy	1323	GCCTAACTTTGTCGCTAGTCTTACAGGAGAAACCTTTAACCAACAAGTTTTTTATCATTTGAA	1382
Db	1381	GCCTAACTTTGTCGCTAGTCTTACAGGAGAAACCTTTAACCAACAAGTTTTTTATCATTTGAA	1440
Db	1383	GACAATATTGAACAACCCCTATTCTTCTGGGATTTGAGAGGGGTGAATAGAGCCTTGAG	1442
Db	1441	GACAATATTGAACAACCCCTATTCTTCTGGGATTTGAGAGGGGTGAATAGAGCCTTGAG	1500
Qy	1443	ACTTTCCCTTTGTGTGTAGGACTTGGAGGAGAAATCCCTTGGACTTTTCACTTAACCCCTCTG	1502
Db	1501	ACTTTCCCTTTGTGTGTAGGACTTGGAGGAGAAATCCCTTGGACTTTTCACTTAACCCCTCTG	1560
Qy	1503	ACATACTCCCAACACCCAGTTGATGGCTTTCCGTAATAAAGANTGGGATT	1554
Db	1561	ACATACTCCCAACACCCAGTTGATGGCTTTCCGTAATAAAGANTGGGATT	1612

RESULT 7

US-902-853-206
; Sequence 206, Application US/09902853
; Publication No. US2002019259A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David

APPLICANT: Botstein, David

88.18; Score 1370; DB 9; Length 1620;

```

: APPLICANT: Desnoyers, Luc
: APPLICANT: Eaton, Dan L.
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, A.
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth, J.
: APPLICANT: Kljavin, Ivar J.
: APPLICANT: Mather, Jennie P.
: APPLICANT: Pan, James
: APPLICANT: Paoni, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William, I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: 10466-14
: CURRENT APPLICATION NUMBER: US/09/902,853
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: US/09/665,350
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US 60/143,048
: PRIOR FILING DATE: 1999-07-07
: PRIOR APPLICATION NUMBER: US 60/145,698
: PRIOR FILING DATE: 1999-07-26
: PRIOR APPLICATION NUMBER: US 60/146,222
: PRIOR FILING DATE: 1999-07-28
: PRIOR APPLICATION NUMBER: PCT/US99/20594
: PRIOR FILING DATE: 1999-09-08
: PRIOR APPLICATION NUMBER: PCT/US99/20944
: PRIOR FILING DATE: 1999-09-13
: PRIOR APPLICATION NUMBER: PCT/US99/21090
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/21547
: PRIOR FILING DATE: 1999-09-15
: PRIOR APPLICATION NUMBER: PCT/US99/23089
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: PCT/US99/28214
: PRIOR FILING DATE: 1999-11-29
: PRIOR APPLICATION NUMBER: PCT/US99/28313
: PRIOR FILING DATE: 1999-11-30
: PRIOR APPLICATION NUMBER: PCT/US99/28564
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/28565
: PRIOR FILING DATE: 1999-12-02
: PRIOR APPLICATION NUMBER: PCT/US99/30095
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: PCT/US99/30911
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US99/30999
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: PCT/US00/00219
: PRIOR FILING DATE: 2000-01-05
: NUMBER OF SEQ ID NOS: 423
: SEQ ID NO 206
: LENGTH: 1620
: TYPE: DNA
: ORGANISM: Homo Sapien
: FEATURE:
: NAME/KEY: unsure
: LOCATION: 973, 977, 996, 1003
: OTHER INFORMATION: unknown base
: US-09-902-853-206

```

Query Match

QY 663 GGCTGCTCATGGACCTTCTCAGGAGAAATGATCCGAGATTTAACTTAATGAGC 722
DB 721 GGCTGCTCATGGACCTTCTCAGGAGAAATGATCCGAGATTTAACTTAATGAGC 780
QY 723 TATACAGCGGGGCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTG 782
DB 781 TATACAGCGGGGCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTG 840
QY 783 TGGCTTCAACCCCAACACAGTGTGAGATGGGGAACACAGAGGATTAATAAGATCCTC 842
DB 841 TGGCTTCAACCCCAACACAGTGTGAGATGGGGAACACAGAGGATTAATAAGATCCTC 900
QY 843 ACTTTGGCAGTGTCTCTCTCTCTCTCAATCCAGGCTTTTCCATACCAACCAACCTGAG 902
DB 901 ACTTTGGCAGTGTCTCTCTCTCTCTCAATCCAGGCTTTTCCATACCAACCAACCTGAG 960
QY 903 GCTGCAGCCTTTTATTTATTTTCCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCAGCT 962
DB 961 GCTGCAGCCTTTTATTTATTTTCCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCAGCT 1020
QY 963 TCTGATTTTAAAGAGGATCTAGGAAATGTGAGGACCCCTACAGGAAGGCTGCCATGC 1022
DB 1021 TCTGATTTTAAAGAGGATCTAGGAAATGTGAGGACCCCTACAGGAAGGCTGCCATGC 1080
1023 TGTGGCCAACTGTTTCACTGAGCAACAAAGAGATCTCATAGGACGAGGGGGAATGGT 1082
DB 1081 TGTGGCCAACTGTTTCACTGAGCAACAAAGAGATCTCATAGGACGAGGGGGAATGGT 1140
QY 1083 TTCCCTCCAAAGCTTGGTCACTGTTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 1142
DB 1141 TTCCCTCCAAAGCTTGGTCACTGTTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 1200
QY 1143 TTTCTCCATGAAGTCTGCTGTTTTCATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1202
DB 1201 TTTCTCCATGAAGTCTGCTGTTTTCATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1260
QY 1203 AGACCTAGATTTTAACTTAAAGTAAAGATGCTGGGATAGAACCTTAAGAAATTTTCCCTC 1262
DB 1261 AGACCTAGATTTTAACTTAAAGTAAAGATGCTGGGATAGAACCTTAAGAAATTTTCCCTC 1320
QY 1263 AAGGACTTCTGCTTCTTAAAGCCTTCTGCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT 1322
DB 1321 AAGGACTTCTGCTTCTTAAAGCCTTCTGCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT 1380
QY 1323 GCTTAAGTCTGCTGCTTCTTAAAGGAGAACCTTTAAACCAACAAAGTTTATCATGATGAA 1382
DB 1381 GCCTAAGTCTGCTGCTTCTTAAAGGAGAACCTTTAAACCAACAAAGTTTATCATGATGAA 1440
QY 1383 GACAATATTGAACAAACCCCTATTTTGTGGGATTTGAGAGGGGTGAATAGAGGCTTGTAG 1442
DB 1441 GACAATATTGAACAAACCCCTATTTTGTGGGATTTGAGAGGGGTGAATAGAGGCTTGTAG 1500
1443 ACTTTCTTCTGCTGCTTCTTAAAGGAGAACCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT 1502
DB 1501 ACTTTCTTCTGCTGCTTCTTAAAGGAGAACCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT 1560
QY 1503 ACATATCTCCCAACCCAGTGTGATGCTTCCGTAATAAAGATTTGGGATTT 1554
DB 1561 ACATATCTCCCAACCCAGTGTGATGCTTCCGTAATAAAGATTTGGGATTT 1612

RESULT 10

US-09-904-011-206
; Sequence 206, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,011
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 206
LENGTH: 1620
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 973, 977, 996, 1003
OTHER INFORMATION: unknown base
US-09-904-011-206

Query Match 88.1%; Score 1370; DB 9; Length 1620;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1494; Conservative 0; Mismatches 4; Indels 114; Gaps 1;

Qy	57	AGATGGCGGTCTTGGCACTCTAATTCGCTCTCGTGTATTCGGTGCCGCGACTTTCACGAT	116
Db	1	AGATGGCGGTCTGGCACTCTAATTCGCTCTCGTGTATTCGGTGCCGCGACTTTCACGAT	60
Qy	117	GGCTCGCCCAACCTTACTACTCTGTGTCGGCCCTGCTCTCTGCTGCGCTTCTCTACTCTCGTA	176
Db	61	GGCTCGCCCAACCTTACTACTCTGTGTCGGCCCTGCTCTCTGCTGCGCTTCTCTACTCTCGTA	120
Qy	177	GGAACCTGCGCGCGCTCTGCGAGGCTCTGCCACCCCAACGCGAAGCGGTAAACCGGTG	236
Db	121	GGAACCTGCGCGCGCTCTGCGAGGCTCTGCCACCCCAACGCGAAGCGGTAAACCGGTG	180
Qy	237	ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCTAGTGCCATTGTCATGATGAAGA	296
Db	181	ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCTAGTGCCATTGTCATGATGAAGA	240
Qy	297	ACCGCAGATCCA-----	308
Db	241	ACCGCAGATCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGG	300
Qy	309	-----	308
Db	301	CCAACAATTCCTTTCTCCGCTTGGATATTCGCATGGGCTACTTTACATCACACTCT	360
Qy	309	-----TGTTCCTGATGAGTGGCAACCCGCCCTATATATGCGGCCCTGAGTATATCAAGT	362
Db	361	GCATAGTGTTCCTGATGACGTGCAACCCGCCCTATATATGCGGCCCTGAGTATATCAAGT	420
Qy	363	ACTTCAATGATAAAACCATTTGATGAGGAACCTAGAAACGGGACCAAGAGGTCACTTGGATTG	422
Db	421	ACTTCAATGATAAAACCATTTGATGAGGAACCTAGAAACGGGACCAAGAGGTCACTTGGATTG	480
Qy	423	TGGAGTCTTTTGCCAAATTTGGTCTTAATGACTGCCAATCATTTGGCCCTATCTATGCTGACC	482
Db	481	TGGAGTCTTTTGCCAAATTTGGTCTTAATGACTGCCAATCATTTGGCCCTATCTATGCTGACC	540
Qy	483	TCTCCCTTAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGAGCGCTATA	542
Db	541	TCCTCCCTTAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGAGCGCTATA	600
Qy	543	CTGATGTTAGTAGCGCGGTACAAAGTAGACACATCACCCCTCACCAAGCAACTCCCTACCC	602
Db	601	CTGATGTTAGTAGCGCGGTACAAAGTAGACACATCACCCCTCACCAAGCAACTCCCTACCC	660
Qy	603	TGATCTCTGTTCCAAAGTGGCAAGAGGCAATGCGGCGGCCACAGATTGCAAGAAAGGAC	662
Db	661	TGATCTCTGTTCCAAAGTGGCAAGAGGCAATGCGGCGGCCACAGATTGCAAGAAAGGAC	720
Qy	663	GGCGTGTCTATGGACCTTCTCTGAGCAGAAATGTGATCCGAGATTTAACTTAATGAGC	722
Db	721	GGCGTGTCTATGGACCTTCTCTGAGCAGAAATGTGATCCGAGATTTAACTTAATGAGC	780
Qy	723	TATACCAAGCGGGCAAGAAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCGCTG	782
Db	781	TATACCAAGCGGGCAAGAAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCGCTG	840
Qy	783	TGGCTTCAACCCCAACACAGTGTGATGGGGAAAAACAAGAGGATAAATAAGATCCTC	842
Db	841	TGGCTTCAACCCCAACACAGTGTGATGGGGAAAAACAAGAGGATAAATAAGATCCTC	900
Qy	843	ACTTTGGCAGTGTCTCTCTCTCTCAATTCACGGCTCTTTCACATACCAACAGCCCTGAG	902
Db	901	ACTTTGGCAGTGTCTCTCTCTCTCAATTCACGGCTCTTTCACATACCAACAGCCCTGAG	960
Qy	903	GCTTCAGCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCAAGCT	962
Db	961	GCTTCAGCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCAAGCT	1020
Qy	963	TCTGATTTTAAAGAGGCATCTAGGGAATTTGTACGGCACCTCTACAGGAAGGCCTGCCATGC	1022
Db	1021	TCTGATTTTAAAGAGGCATCTAGGGAATTTGTACGGCACCTCTACAGGAAGGCCTGCCATGC	1080

RESULT 11

US-10-174-590-31
; Sequence 31, Application US/10174590

: PUBLICATION NO: 0320
: GENERAL INFORMATION:

APPLICANT: Baker, Kevin P

• APPLICANT: Chen, Tian

APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Au:

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBR

; TITLE OF INVENTION: ACIDS ENCODING THE S

FILE REFERENCE: P3430R1C42

; CURRENT APPLICATION NUMBER: US/10/174,590

1 CURRENT FILLING

CONSENT FILING : Prior applica

NUMBER OF SECTIONS

; NUMBER OF SEQ ID NOS:
 : SEQ ID NO 31

; SEQ ID NO
: LENGTH.

```

; LENGTH: 1620
; TYPE: DNA

```

TYPE: DNA
ORGANISM: Homo sapiens

; ORGANISM: Homo sapien
 ; FEATURE:

```

; FEATURE:
; NAME / KEY:

```

Query Match

88.1%: score 1370: DA 9: Length 1620:

Query Match
Best Local Similarity
08.1%
Score 1370
DB 3
Length 1020
Pred No 0
92.7%

BEST LOCAL SIMILARITY 92.7%; PRED: NO; 0;
Matches 1494: Conservative 0: Mismatches 4: Indels 114: Cases 1:

Mismatches 1494; Conservative 0; Mismatches 4; Indels 114; Gaps 17;

QY 57 AGATGGCGGTCTTGGCACCTCTAATTCCTCTCTGTATTCGGTGGCCGACCTTTCAGCAT 116
DB 1 AGATGGCGGTCTTGGCACCTCTAATTCCTCTCTGTATTCGGTGGCCGACCTTTCAGCAT 60
QY 117 GGTCTGCCCAACCTTACTTCTCTGTGCGCCCTGTCTCTGTCTGCTTCTTCTACTCTGTGA 176
DB 61 GGTCTGCCCAACCTTACTTCTCTGTGCGCCCTGTCTCTGTCTGCTTCTTCTACTCTGTGA 120
QY 177 GGAATCTGCCCGGCTCTGCCAGGGTCTGCCACCCCAACGCAAGCGGTAACCCCGTGTG 236
DB 121 GGAATCTGCCCGGCTCTGCCAGGGTCTGCCACCCCAACGCAAGCGGTAACCCCGTGTG 180
QY 237 ACTTTGACTGGAGAGAGTGGAGATCCTGTATGTTTCTCAGTGCATTTGATGATGAGA 296
DB 181 ACTTTGACTGGAGAGAGTGGAGATCCTGTATGTTTCTCAGTGCATTTGATGATGAGA 240
QY 297 ACCGAGATGCCA----- 308
DB 241 ACCGAGATGCCA----- 300
QY 309 ----- 308
DB 301 CCAACACAATTTCTTCTCGCTTGGATATTCGCATGGCCCTTACATCACACTCT 360
QY 309 -----TGTCTCTGATGAGTGCACACCCCTCTATATATGCGCCCTGAGTATATCAAGT 362
DB 361 GCATAGTGTCTCTGATGAGTGCACACCCCTCTATATATGCGCCCTGAGTATATCAAGT 420
QY 363 ACTTCAATGATAAACCATTGATGAGGAAGTACAGCGGACAGAGGGTCACTTGGATTG 422
DB 421 ACTTCAATGATAAACCATTGATGAGGAAGTACAGCGGACAGAGGGTCACTTGGATTG 480
QY 423 TGGAGTCTTTGCAATTTGCTTAATGCTTAATGACTGCGCAATCATTTGCCCTCTATCTATGCTGACC 482
DB 481 TGGAGTCTTTGCAATTTGCTTAATGCTTAATGACTGCGCAATCATTTGCCCTCTATCTATGCTGACC 540
QY 483 TCTCCCTTAATACACTGTACAGGGCTAAATTTTGGGAAGGTGGATTTGGAGCGCTATA 542
DB 541 TCTCCCTTAATACACTGTACAGGGCTAAATTTTGGGAAGGTGGATTTGGAGCGCTATA 600
QY 543 CTGATCTTAGTCGGGTACAAAGTGCAGCATACCCCTCACCAAGCACTCCCTACCC 602
DB 601 CTGATCTTAGTCGGGTACAAAGTGCAGCATACCCCTCACCAAGCACTCCCTACCC 660
QY 603 TGATCTCTTCCAAAGTGCAGGAGGCAATGCGCGGCCACAGATTGCAAGAAAGGAC 662
DB 661 TGATCTCTTCCAAAGTGCAGGAGGCAATGCGCGGCCACAGATTGCAAGAAAGGAC 720
QY 663 GGGCTGCTCATGGACCTTCTCTGAGGAGAAATGTATCGGAGAAATTTAACTTAATGAGC 722
DB 721 GGGCTGCTCATGGACCTTCTCTGAGGAGAAATGTATCGGAGAAATTTAACTTAATGAGC 780
QY 723 TATACAGCGGGCCAAAGAACTATCAAGGCTGGAGCAATATCCCTGAGGAGCAGCCTG 782
DB 781 TATACAGCGGGCCAAAGAACTATCAAGGCTGGAGCAATATCCCTGAGGAGCAGCCTG 840
QY 783 TGGCTTCAACCCCAACAGTGTCTGAGTGGGAAACAAAGAGGATAAATAAGATCCCTC 842
DB 841 TGGCTTCAACCCCAACAGTGTCTGAGTGGGAAACAAAGAGGATAAATAAGATCCCTC 900
QY 843 ACTTTGGCAGTGTCTCTCTCTCTCAATTCAGGCTCTTTCATACACACAGCCTGAG 902
DB 901 ACTTTGGCAGTGTCTCTCTCTCTCAATTCAGGCTCTTTCATACACACAGCCTGAG 960
QY 903 GCTGACGCTTTTATTTATTTTCTCTTCTGCTGTGCTGGGTGGGCGAGCATGCAGCT 962
DB 961 GCTGACGCTTTTATTTATTTTCTCTTCTGCTGTGCTGGGTGGGCGAGCATGCAGCT 1020
QY 963 TCTGATTTTAAAGAGCATCTAGGGAATTTGTCAGGCAACCCCTACAGGAAGCCTTGCATGC 1022
DB 1021 TCTGATTTTAAAGAGCATCTAGGGAATTTGTCAGGCAACCCCTACAGGAAGCCTTGCATGC 1080

QY 1023 TGTGGCCAACTGTTTCACTGGAGCAAGAGAGATCTCATAGGACGAGGGGAAATGGT 1082
DB 1081 TGTGGCCAACTGTTTCACTGGAGCAAGAGAGATCTCATAGGACGAGGGGAAATGGT 1140
QY 1083 TTTCCCTCCAAAGCTTGGGTCAAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 1142
DB 1141 TTTCCCTCCAAAGCTTGGGTCAAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 1200
QY 1143 TTTCTCCATGAAGAACTCTGTGGTTTTCATCATCTCTTCTTATGACCTGACAGCTTGGTT 1202
DB 1201 TTTCTCCATGAAGAACTCTGTGGTTTTCATCATCTCTTCTTATGACCTGACAGCTTGGTT 1260
QY 1203 AGACCTAGATTTAAACCTTAAGGTAAGATGCTGGGTATAGAAGCTAAGAAATTTTCCCCC 1262
DB 1261 AGACCTAGATTTAAACCTTAAGGTAAGATGCTGGGTATAGAAGCTAAGAAATTTTCCCCC 1320
QY 1263 AAGGACTCTTGTCTTAAAGCCCTTCTGCTTCTGTTTATGGTCTTCAATTAAGATATAA 1322
DB 1321 AAGGACTCTTGTCTTAAAGCCCTTCTGCTTCTGTTTATGGTCTTCAATTAAGATATAA 1380
QY 1323 GCCTAATTTGTGCTAGTCTTAAGGAGAAACCTTTAAACCAAAAGTTTATCATATTGAA 1382
DB 1381 GCCTAATTTGTGCTAGTCTTAAGGAGAAACCTTTAAACCAAAAGTTTATCATATTGAA 1440
QY 1383 GACAATATTCAACAAACCCCTATTTTGTGGGATTTGAGAGGGGTGAATAGAGCTTGAG 1442
DB 1441 GACAATATTCAACAAACCCCTATTTTGTGGGATTTGAGAGGGGTGAATAGAGCTTGAG 1500
QY 1443 ACTTCTCTTGTGTGTAGGACTTGGAGGAGAAATCCCTCGGACTTTTCACTAACCCCTCTG 1502
DB 1501 ACTTCTCTTGTGTGTAGGACTTGGAGGAGAAATCCCTCGGACTTTTCACTAACCCCTCTG 1560
QY 1503 ACATCTCCCAACCCAGTTGATGCTTTCCGTAATAAAGATTTGGGATT 1554
DB 1561 ACATCTCCCAACCCAGTTGATGCTTTCCGTAATAAAGATTTGGGATT 1612

RESULT 14

US-09-906-742-206
; Sequence 206, Application US/09906742
; Publication No. US20030023054A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,742
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350

```

; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 206
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 973, 977, 996, 1003
; OTHER INFORMATION: unknown base
09-906-742-206

Query Match      88.1%; Score 1370; DB 9; Length 1620;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1494; Conservative 0; Mismatches 4; Indels 114; Gaps 1;

Oy 57 AGATGGCGTCTGTGACCTCTAAATGCTCTCGTGATTCGGTGGCGGACATTCACGAT 116
Db 1 AGATGGCGTCTGTGACCTCTAAATGCTCTCGTGATTCGGTGGCGGACATTCACGAT 60

Oy 117 GGCTGCGCCCAACTTACTACCTCTGTGCGGCCCTGCTCTGCTGCTTCTACTCGTGA 176
Db 61 GGCTGCGCCCAACTTACTACCTCTGTGCGGCCCTGCTCTGCTGCTTCTACTCGTGA 120

Oy 177 GAAACTGCGCGGCTCTGCCAGGTCTGCCACCCCAACGCGGAAGACGGTAACCCGTGTG 236
Db 121 GAAACTGCGCGGCTCTGCCAGGTCTGCCACCCCAACGCGGAAGACGGTAACCCGTGTG 180

Oy 237 ACTTFTGCTGGAGAGAAGTGGAGATCCCTGATGTTTCTCAGTGCCCATTTGTGATGAAGA 296
Db 181 ACTTFTGCTGGAGAGAAGTGGAGATCCCTGATGTTTCTCAGTGCCCATTTGTGATGAAGA 240

Oy 297 ACCGAGATCCA----- 308
Db 241 ACCGAGATCCA----- 300

Oy 309 ----- 308
Db 301 CCAACAAATCTTTTCTCCGCTTGGGATATTCGGATGGGCTACTTTTACATCACACTCT 360

309 -----TGTTCCTGATGACGTGCAAAACCCCTATATATATGCGCCTGAGTATATCAAGT 362
361 GCATAGTGTTCCTGATGACGTGCAAAACCCCTATATATATGCGCCTGAGTATATCAAGT 420
363 ACTTCAATGATAAAACCATTTGATGAGAACTAGAACGGGACAAAGAGGTCACATTGGATTG 422
421 ACTTCAATGATAAAACCATTTGATGAGAACTAGAACGGGACAAAGAGGTCACATTGGATTG 480
423 TGGAGTTCCTTGGCAATTTGGTCTAATGACTGCCAAATCATTTGCCCTTATCTATGCTGACC 482
481 TGGAGTTCCTTGGCAATTTGGTCTAATGACTGCCAAATCATTTGCCCTTATCTATGCTGACC 540
483 TCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGAAAGGTGGATTTGGAGGCTATA 542
541 TCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGAAAGGTGGATTTGGAGGCTATA 600
543 CTGATGTTAGTAGCGGTACAAAGTAGGACACATCACCCCTCACCAAGCAACTCCCTACCC 602
601 CTGATGTTAGTAGCGGTACAAAGTAGGACACATCACCCCTCACCAAGCAACTCCCTACCC 660
603 TGATCCTGTTCCAAAGTGGCAAGGAGCAATGCGGGGCCACAGATTGACAAAGAGGAC 662
661 TGATCCTGTTCCAAAGTGGCAAGGAGCAATGCGGGGCCACAGATTGACAAAGAGGAC 720
663 GGGCTGCTCATGGACCTTCTCTGAGGAGAAATGTCGAGAAATTAATTAATGAGC 722
721 GGGCTGCTCATGGACCTTCTCTGAGGAGAAATGTCGAGAAATTAATTAATGAGC 780
723 TATACCAGGGGCAAGAACTATCAAGGCTGGAGACAAATATCCCTGAGGACGACGCTG 782
781 TATACCAGGGGCAAGAACTATCAAGGCTGGAGACAAATATCCCTGAGGACGACGCTG 840
783 TGGCTTCAACCCGCCACAGTGTGATGGGAAACAAAGAGAGTAAATAAGATCCTC 842
841 TGGCTTCAACCCGCCACAGTGTGATGGGAAACAAAGAGAGTAAATAAGATCCTC 900
843 ACTTTGGCAGTGTCTCTCTCTGATTCAGGCTCTTCCATACCAACAGCCTGAG 902
901 ACTTTGGCAGTGTCTCTCTCTCTGATTCAGGCTCTTCCATACCAACAGCCTGAG 960
903 GCTGACGCTTTTATTTATGTTTCCCTTGTGGCTGTGACTGGTGGGCGGACGATGACGT 962
961 GCTGACGCTTTTATTTATGTTTCCCTTGTGGCTGTGACTGGTGGGCGGACGATGACGT 1020
963 TCTGATTTTAAAGAGGACATCTAGGGAATGTCAGGCACCCCTACAGGAAGCCCTGCCATGC 1022
1021 TCTGATTTTAAAGAGGACATCTAGGGAATGTCAGGCACCCCTACAGGAAGCCCTGCCATGC 1080
1023 TGTGGCCAACTGTTTCACTGGAGCAAGAAAGATCTCATAGGACGGAGGGGAAATGGT 1082
1081 TGTGGCCAACTGTTTCACTGGAGCAAGAAAGATCTCATAGGACGGAGGGGAAATGGT 1140
1083 TTCCCTCCAACTGTTGGTCTGAGTGTAACTGCTTATCAGTATTCAGACATCTCCATGG 1142
1141 TTCCCTCCAACTGTTGGTCTGAGTGTAACTGCTTATCAGTATTCAGACATCTCCATGG 1200
1143 TTTCTCCATGAACTCTGTGGTTTTCATCTTCTTCTAGTTGACCTGCACACGCTTGGTT 1202
1201 TTTCTCCATGAACTCTGTGGTTTTCATCTTCTTCTAGTTGACCTGCACACGCTTGGTT 1260
1203 AGACCTAGATTTAACCCCTAAGGTAAGATGCTGGGGTATAGACGCTAAGAAATTTTCCCC 1262
1261 AGACCTAGATTTAACCCCTAAGGTAAGATGCTGGGGTATAGACGCTAAGAAATTTTCCCC 1320
1263 AGGAGCTCTTGTCTTAAAGCCCTTCTGCTTCTGCTTCTTATGCTTCTTAAAGATATA 1322
1321 AGGAGCTCTTGTCTTAAAGCCCTTCTGCTTCTGCTTCTTATGCTTCTTAAAGATATA 1380
1323 GCTTAACTTTGTGCTAGTCTTAAAGGAGAAACCTTTAAACCAACAAAGCTTTTATCATTTGAA 1382
1381 GCCTAACTTTGTGCTAGTCTTAAAGGAGAAACCTTTAAACCAACAAAGCTTTTATCATTTGAA 1440
```

```
Qy 1383 GACAATATTGAACACCCCTATTGTTGGGATTGAGAAGGGTGAATAGAGGCTTGAG 1442
Db 1441 GACANTATTGAACACCCCTATTGTTGGGATTGAGAAGGGTGAATAGAGGCTTGAG 1500
Qy 1443 ACTTTCCTTTGTGTGGTAGGACTTTGGAGGAGAAATCCCTGGACTTTTCACATAACCCCTCTG 1502
Db 1501 ACTTTCCTTTGTGTGGTAGGACTTTGGAGGAGAAATCCCTGGACTTTTCACATAACCCCTCTG 1560
Qy 1503 ACATACCTCCCAACACCCAGTTGATGGCTTTCGGTAAATAAAGATTGGGATT 1554
Db 1561 ACATACCTCCCAACACCCAGTTGATGGCTTTCGGTAAATAAAGATTGGGATT 1612

RESULT 15
US-10-173-706-31
; Sequence 31, Application US/10173706
; Publication No. US2003002293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 31
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 973..977, 996, 1003
; OTHER INFORMATION: unknown base
;
US-10-173-706-31
Query Match 88.1%; Score 1370; DB 9; Length 1620;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1494; Conservative 0; Mismatches 4; Indels 114; Gaps 1;

Qy 57 AGATGGCGGTCTTGACACCTCTAATTGCTCTCGTGTATTCGGTGGCGGACTTTTCACGAT 116
1 AGATGGCGGTCTTGACACCTCTAATTGCTCTCGTGTATTCGGTGGCGGACTTTTCACGAT 60
Qy 117 GGCTCGCCCAACCTTACTTACTCTCTCGCGCCCTGCTCTGCTGCTTCTACTCTGTA 176
61 GGCTCGCCCAACCTTACTTACTCTCTCGCGCCCTGCTCTGCTGCTTCTACTCTGTA 120
Qy 177 GGAAACTCGCCCGCTCTGCGCACGCTCTGCGCACCCCAACCGGAGACGGTAACCGGTGTG 236
121 GGAAACTCGCCCGCTCTGCGCACGCTCTGCGCACCCCAACCGGAGACGGTAACCGGTGTG 180
Qy 237 ACTTTGACTGGAGAGAGTGAGATCCCTGATGTTTCTCAGTGCCCATTTGATGATGAAGA 296
181 ACTTTGACTGGAGAGAGTGAGATCCCTGATGTTTCTCAGTGCCCATTTGATGATGAAGA 240
Qy 297 ACCGAGATCCA----- 308
241 ACCGAGATCCATCATCTGAGGAGCAACATATAGGCAACATTTTCATGTTTAAAGTGG 300
309 ----- 308
301 CCAACACAAATCTTTTCTCCGCTGGGATATTCGGATGGGCTACTTTACATCACACTCT 360
```

```
Qy 309 -----TGTTCTCTGATGACGTGCAAAACCCCTATATATATGGCCCTCAGTATATCAAGT 362
Db 361 GCATAGTGTCTCTGATGACGTGCAAAACCCCTATATATATGGCCCTCAGTATATCAAGT 420
Qy 363 ACTTCAATGATAAAACCAATTGATGAGGAACCTAGAACGGGACAAAGAGGTCTACTTGGATTG 422
Db 421 ACTTCAATGATAAAACCAATTGATGAGGAACCTAGAACGGGACAAAGAGGTCTACTTGGATTG 480
Qy 423 TGGAGTCTTTGGCCAAATTGGTCTAATGACTGCCAATCATTTGCCCTCTATCTATGCTGACC 482
Db 481 TGGAGTCTTTGGCCAAATTGGTCTAATGACTGCCAATCATTTGCCCTCTATCTATGCTGACC 540
Qy 483 TCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGATGTTGGACGCTATA 542
Db 541 TCTCCCTTAAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGATGTTGGACGCTATA 600
Qy 543 CTGATGTTAGTACGCGGTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCC 602
Db 601 CTGATGTTAGTACGCGGTACAAAGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCC 660
Qy 603 TGATCCTGTTCOAAGGTGGCAAGAGGCAATCGCGGGCCACAGATTGACAAGAAAGGAC 662
Db 661 TGATCCTGTTCOAAGGTGGCAAGAGGCAATCGCGGGCCACAGATTGACAAGAAAGGAC 720
Qy 663 GGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCCGAGAAATTTAAATGAGC 722
Db 721 GGGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCCGAGAAATTTAAATGAGC 780
Qy 723 TATACACGCGGGCCAAAGAACTATCAAAAGGCTGGAGACAATATCCCTCAGGAGCAGCCTG 782
Db 781 TATACACGCGGGCCAAAGAACTATCAAAAGGCTGGAGACAATATCCCTCAGGAGCAGCCTG 840
Qy 783 TGGCTTCAACCCCCACCACAGTGTGATGGGGAAGAAAGAGGATAAATAGATCTCTC 842
Db 841 TGGCTTCAACCCCCACCACAGTGTGATGGGGAAGAAAGAGGATAAATAGATCTCTC 900
Qy 843 ACTTTGGCAGTGTCTCTCTCTCAATTCCAGGCTCTTTCCATACCCACCAAGCCTGAG 902
Db 901 ACTTTGGCAGTGTCTCTCTCTCAATTCCAGGCTCTTTCCATACCCACCAAGCCTGAG 960
Qy 903 GCTGAGCCTTTTATTTATGTTTCTCTCTGCTGTGCTGGTGGGGCAGCATGCAGCT 962
Db 961 GCTGAGCCTTTTATTTATGTTTCTCTCTGCTGTGCTGGTGGGGCAGCATGCAGCT 1020
Qy 963 TCTGATTTTAAAGAGGATCTAGGGAATGTGAGGACCCCTACAGGAAGCCCTGCCATGC 1022
Db 1021 TCTGATTTTAAAGAGGATCTAGGGAATGTGAGGACCCCTACAGGAAGCCCTGCCATGC 1080
Qy 1023 TGTGGCCCAACTGTTTCACTGGGAGCAAGAGAGATCTCATAGGAGGGGGAATGCT 1082
Db 1081 TGTGGCCCAACTGTTTCACTGGGAGCAAGAGAGATCTCATAGGAGGGGGAATGCT 1140
Qy 1083 TTCCCTCCAAAGCTGGGTCAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATGG 1142
Db 1141 TTCCCTCCAAAGCTGGGTCAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATGG 1200
Qy 1143 TTTCCTCATGAAACTCTGTGGTTTTCATCATCTCTTTAGTTGAACTGCACAGCTGGTT 1202
Db 1201 TTTCCTCATGAAACTCTGTGGTTTTCATCATCTCTTTAGTTGAACTGCACAGCTGGTT 1260
Qy 1203 AGACCTAGATTTAAACCTTAAGTATAGATCTGGGTATAGACGCTAAGAAATTTTCCCCC 1262
Db 1261 AGACCTAGATTTAAACCTTAAGTATAGATCTGGGTATAGACGCTAAGAAATTTTCCCCC 1320
Qy 1263 AAGGACTCTTGTCTTCCCTTAAAGCCCTTCTGGCTTATGCTTCTTAAAGATATAA 1322
Db 1321 AAGGACTCTTGTCTTCCCTTAAAGCCCTTCTGGCTTATGCTTCTTAAAGATATAA 1380
Qy 1323 GCCTAACTTTGTGCTAGTCTCTTAAGGAGAAACCTTTAAACCAACAAAGTTTTTATCATTTGAA 1382
Db 1381 GCCTAACTTTGTGCTAGTCTCTTAAGGAGAAACCTTTAAACCAACAAAGTTTTTATCATTTGAA 1440
```

QY 1383 GACAATATTGAACACCCCTATTCTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAG 1442
|||||
Db 1441 GACAATATTGAACACCCCTATTCTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAG 1500
|||||
QY 1443 ACTTTCCTTTGTGTAGGACTTGGAGGAGAAATCCCTGGACTTTCACCTAACCCCTCTG 1502
|||||
Db 1501 ACTTTCCTTTGTGTAGGACTTGGAGGAGAAATCCCTGGACTTTCACCTAACCCCTCTG 1560
|||||
QY 1503 ACATACTCCCCACACCCAGTTGATGGCTTTCCTAATAAAAAAGATTGGGATT 1554
|||||
Db 1561 ACATACTCCCCACACCCAGTTGATGGCTTTCCTAATAAAAAAGATTGGGATT 1612
|||||

Search completed: July 12, 2003, 11:10:04
Job time : 218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 11:54:20 ; Search time 23 Seconds
(without alignments)
1305.926 Million cell updates/sec

Title: US-09-954-846-2
Perfect score: 1359
Sequence: 1 MAVLAPLIALVYSPRLSRW.....EQVASTPTTVSDGENKDK 258

Scoring table: BLOSUM62
Gapex 10.0, Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1359	100.0	258	10	US-09-954-846-2 Sequence 2, Appl
2	1359	100.0	272	10	US-09-925-300-1691 Sequence 1691, Ap
3	1326	97.6	296	9	US-09-905-291A-207 Sequence 207, App
4	1326	97.6	296	9	US-09-902-853-207 Sequence 207, App
5	1326	97.6	296	9	US-09-907-824-207 Sequence 207, App
6	1326	97.6	296	9	US-09-907-841-207 Sequence 207, App
7	1326	97.6	296	9	US-09-904-011-207 Sequence 207, App
8	1326	97.6	296	9	US-10-174-590-32 Sequence 32, Appl
9	1326	97.6	296	9	US-10-176-758-32 Sequence 32, Appl
10	1326	97.6	296	9	US-10-175-737-32 Sequence 32, Appl
11	1326	97.6	296	9	US-09-906-742-207 Sequence 207, App
12	1326	97.6	296	9	US-10-173-706-32 Sequence 32, Appl
13	1326	97.6	296	9	US-10-175-738-32 Sequence 32, Appl
14	1326	97.6	296	9	US-10-175-752-32 Sequence 32, Appl
15	1326	97.6	296	9	US-10-176-482-32 Sequence 32, Appl
16	1326	97.6	296	9	US-10-176-757-32 Sequence 32, Appl
17	1326	97.6	296	9	US-10-176-913-32 Sequence 32, Appl
18	1326	97.6	296	9	US-10-180-552-32 Sequence 32, Appl
19	1326	97.6	296	9	US-10-180-557-32 Sequence 32, Appl

20	1326	97.6	296	9	US-09-906-838-207 Sequence 207, App
21	1326	97.6	296	9	US-09-907-613-207 Sequence 207, App
22	1326	97.6	296	9	US-09-907-942-207 Sequence 207, App
23	1326	97.6	296	9	US-10-173-700-32 Sequence 32, Appl
24	1326	97.6	296	9	US-10-174-572-32 Sequence 32, Appl
25	1326	97.6	296	9	US-10-174-579-32 Sequence 32, Appl
26	1326	97.6	296	9	US-10-174-582-32 Sequence 32, Appl
27	1326	97.6	296	9	US-10-174-588-32 Sequence 32, Appl
28	1326	97.6	296	9	US-10-175-739-32 Sequence 32, Appl
29	1326	97.6	296	9	US-10-175-740-32 Sequence 32, Appl
30	1326	97.6	296	9	US-10-175-743-32 Sequence 32, Appl
31	1326	97.6	296	9	US-10-176-488-32 Sequence 32, Appl
32	1326	97.6	296	9	US-10-176-492-32 Sequence 32, Appl
33	1326	97.6	296	9	US-10-176-747-32 Sequence 32, Appl
34	1326	97.6	296	9	US-10-176-750-32 Sequence 32, Appl
35	1326	97.6	296	9	US-10-176-985-32 Sequence 32, Appl
36	1326	97.6	296	9	US-10-176-987-32 Sequence 32, Appl
37	1326	97.6	296	9	US-10-176-991-32 Sequence 32, Appl
38	1326	97.6	296	9	US-10-176-992-32 Sequence 32, Appl
39	1326	97.6	296	9	US-10-176-993-32 Sequence 32, Appl
40	1326	97.6	296	9	US-10-184-658-32 Sequence 32, Appl
41	1326	97.6	296	9	US-10-173-695-32 Sequence 32, Appl
42	1326	97.6	296	9	US-10-173-697-32 Sequence 32, Appl
43	1326	97.6	296	9	US-10-173-705-32 Sequence 32, Appl
44	1326	97.6	296	9	US-10-174-576-32 Sequence 32, Appl
45	1326	97.6	296	9	US-10-174-585-32 Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-954-846-2
; Sequence 2, Application US/09954846
; Patent No. US20020102654A1
; GENERAL INFORMATION:
; APPLICANT: Tanq, Y. Tom
; Corley, Neil C.
; Guegler, Karl J.
; Patterson, Chandra
; Baughn, Mariah R.
; TITLE OF INVENTION: THIOREDOXIN PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/954,846
; FILING DATE: 17-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/107,248
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0556 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 amino acids

```
;
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT19
; CLONE: 3244141
; SEQUENCE DESCRIPTION: SEQ ID NO: 2 :
US-09-954-846-2

Query Match      100.0%; Score 1359; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 3.6e-133;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMKNRRSMFLMTCKPPLYMGPEYIKYFNDKTIDELELRDKRVTV 120
Db 61 FDWREVEILMFLSAIVMKNRRSMFLMTCKPPLYMGPEYIKYFNDKTIDELELRDKRVTV 120
Qy 121 IVEFFANWSDCOSFAPYIADLSLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKQLP 180
Db 121 IVEFFANWSDCOSFAPYIADLSLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKQLP 180
Qy 181 TLILFOGKKEAMRRPQIDKKGRAVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQ 240
Db 181 TLILFOGKKEAMRRPQIDKKGRAVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQ 240
Qy 241 PVASTPTTVSDGENKKDK 258
Db 241 PVASTPTTVSDGENKKDK 258

RESULT 2
US-09-925-300-1691
; Sequence 1691, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1691
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1691

Query Match      100.0%; Score 1359; DB 10; Length 272;
Best Local Similarity 100.0%; Pred. No. 3.9e-133;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 60
Db 15 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAFLVLRKLPPLCHGLPTQREDGNPCD 74
Qy 61 FDWREVEILMFLSAIVMKNRRSMFLMTCKPPLYMGPEYIKYFNDKTIDELELRDKRVTV 120
Db 75 FDWREVEILMFLSAIVMKNRRSMFLMTCKPPLYMGPEYIKYFNDKTIDELELRDKRVTV 134
Qy 121 IVEFFANWSDCOSFAPYIADLSLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKQLP 180
Db 135 IVEFFANWSDCOSFAPYIADLSLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKQLP 194

Qy 181 TLILFOGKKEAMRRPQIDKKGRAVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQ 240
Db 195 TLILFOGKKEAMRRPQIDKKGRAVSWTFSEENVIREFNELNYORAKKLSKAGDNIPEEQ 254
Qy 241 PVASTPTTVSDGENKKDK 258
Db 255 PVASTPTTVSDGENKKDK 272

RESULT 3
US-09-905-291A-207
; Sequence 207, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pah, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
```

;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 207
;; LENGTH: 296
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-905-291A-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
|||||
1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
|||||
61 FDMREVEILMFLSAIVMKNRRS----- 83
|||||
61 FDMREVEILMFLSAIVMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
|||||
84 -MFLMTCCKPLYMGPEYIKYFNDKTIIDEELERDKRVTWIVEFFANWSNDCOSFAPYADL 142
:|||||
121 IVFLMTCCKPLYMGPEYIKYFNDKTIIDEELERDKRVTWIVEFFANWSNDCOSFAPYADL 180
|||||
143 SLKYNCTGLNFGKVDVGRYTDVSTRYKVSPTLKQPLTLILFQGGKEAMRRPQIDKKGR 202
|||||
181 SLKYNCTGLNFGKVDVGRYTDVSTRYKVSPTLKQPLTLILFQGGKEAMRRPQIDKKGR 240
|||||
203 AVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
|||||
241 AVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
|||||

RESULT 4
US-09-902-853-207
; Sequence 207, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350

;; PRIOR FILING DATE: 2000-09-18
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 207
;; LENGTH: 296
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-902-853-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

Qy 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
|||||
Db 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
|||||
Qy 61 FDMREVEILMFLSAIVMKNRRS----- 83
|||||
Db 61 FDMREVEILMFLSAIVMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
|||||
Qy 84 -MFLMTCCKPLYMGPEYIKYFNDKTIIDEELERDKRVTWIVEFFANWSNDCOSFAPYADL 142
:|||||
Db 121 IVFLMTCCKPLYMGPEYIKYFNDKTIIDEELERDKRVTWIVEFFANWSNDCOSFAPYADL 180
|||||
Qy 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKVSPTLKQPLTLILFQGGKEAMRRPQIDKKGR 202
|||||
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKVSPTLKQPLTLILFQGGKEAMRRPQIDKKGR 240
|||||
Qy 203 AVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
|||||
Db 241 AVSWTFSEENVIREFNLNELYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
|||||

RESULT 5
US-09-907-824-207
; Sequence 207, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: US/09/907,824
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 207
LENGTH: 296
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-824-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSVPRLSRWLAQPYLLSALLSAFLVRLKPLPCHGLPTOREDGNPCD 60
DB 1 MAVLAPLIALVYSVPRLSRWLAQPYLLSALLSAFLVRLKPLPCHGLPTOREDGNPCD 60
QY 61 FDWREVEIIMFLSAIVMMKNRRS-----83
DB 61 FDWREVEIIMFLSAIVMMKNRRSITVEQHIGNIFWFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIIDELEKRDKRVTWIIEFFANMSNDCQSFAPYADL 142
DB 121 IVFLMTCKPPLYMGPEYIKYFNDKTIIDELEKRDKRVTWIIEFFANMSNDCQSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGKEAMRRPQIDKKGR 202
DB 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNLNELYQRAKKLSKAGNIPPEQPVASTPTTVSDGENKKDK 258
DB 241 AVSWTFSEENVIREFNLNELYQRAKKLSKAGNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 6

US-09-907-841-207
Sequence 207, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15

;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 207
;; LENGTH: 296
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-907-841-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

1 MAVLAPLALVSVPRLSRWLAOPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
1 MAVLAPLALVSVPRLSRWLAOPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
61 FQWREVEILMFLSAIVMKNNRS----- 83
61 FQWREVEILMFLSAIVMKNNRSITVEQHIIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
84 -MFLMTCKPPLYMGPEYIKYFNDKTIIDELEDRKRVTWIVEFFANWSNDCQSFAPYIADL 142
121 IVFLMTCKPPLYMGPEYIKYFNDKTIIDELEDRKRVTWIVEFFANWSNDCQSFAPYIADL 180
143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFQGGKEAMRRPOIDKKGR 202
181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFQGGKEAMRRPOIDKKGR 240
203 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
241 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 7
US-09-904-011-207
; Sequence 207, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350

;; PRIOR FILING DATE: 2000-09-18
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 207
;; LENGTH: 296
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-904-011-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLALVSVPRLSRWLAOPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
DB 1 MAVLAPLALVSVPRLSRWLAOPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
QY 61 FQWREVEILMFLSAIVMKNNRS----- 83
DB 61 FQWREVEILMFLSAIVMKNNRSITVEQHIIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIIDELEDRKRVTWIVEFFANWSNDCQSFAPYIADL 142
DB 121 IVFLMTCKPPLYMGPEYIKYFNDKTIIDELEDRKRVTWIVEFFANWSNDCQSFAPYIADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFQGGKEAMRRPOIDKKGR 202
DB 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLILFQGGKEAMRRPOIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
DB 241 AVSWTFSEENVIREFNELNYQRAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 8
US-10-174-590-32
; Sequence 32, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-32

Query Match          97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
QY 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 9
US-10-176-758-32
; Sequence 32, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-32

Query Match          97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
QY 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 9
US-10-176-758-32
; Sequence 32, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-32

Query Match          97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
QY 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
```

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-32

Query Match          97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
QY 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296

RESULT 10
US-10-175-737-32
; Sequence 32, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-32

Query Match          97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVLRKLPPLCHGLPTQREDGNCPCD 60
QY 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNDKTIDELEDRKRVTWIVFFANWSNDCOSFAPYADL 180
QY 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGCKEAMRRPQIDKKGR 240
QY 203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
```

QY 143 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGGEAMRRPQIDKKGR 202
|||||
Db 181 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGGEAMRRPQIDKKGR 240
|||||
QY 203 AVSWTFSEENVIREFNELNELYORAKKLSKAGDNIPPEQVASTPTTVSDGENKKDK 258
|||||
Db 241 AVSWTFSEENVIREFNELNELYORAKKLSKAGDNIPPEQVASTPTTVSDGENKKDK 296
|||||

RESULT 11

US-09-906-742-207
; Sequence 207, Application US/09906742
; Publication No. US20030023054A1

GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kiljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14

; CURRENT APPLICATION NUMBER: US/09/906,742
; CURRENT FILING DATE: 2001-07-16

; PRIOR APPLICATION NUMBER: 09/665,350

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: PCT/US00/04414

; PRIOR FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 60/143,048

; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 60/145,698

; PRIOR FILING DATE: 1999-07-26

; PRIOR APPLICATION NUMBER: US 60/146,222

; PRIOR FILING DATE: 1999-07-28

; PRIOR APPLICATION NUMBER: PCT/US99/20594

; PRIOR FILING DATE: 1999-09-08

; PRIOR APPLICATION NUMBER: PCT/US99/20944

; PRIOR FILING DATE: 1999-09-13

; PRIOR APPLICATION NUMBER: PCT/US99/21090

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/21547

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/23089

; PRIOR FILING DATE: 1999-10-05

; PRIOR APPLICATION NUMBER: PCT/US99/28214

; PRIOR FILING DATE: 1999-11-29

; PRIOR APPLICATION NUMBER: PCT/US99/28313

; PRIOR FILING DATE: 1999-11-30

; PRIOR APPLICATION NUMBER: PCT/US99/28564

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 207
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-742-207

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

QY 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVVRKLPPLCHGLPTQREDGNPCD 60
|||||
Db 1 MAVLAPLIALVYSPRLSRWLAQPYLLSALLSAAFLVVRKLPPLCHGLPTQREDGNPCD 60
|||||
QY 61 FDMREVEILMFLSAIYVMKNRRS----- 83
|||||
Db 61 FDMREVEILMFLSAIYVMKNRRSITVEQHIGNIFMFSKVANTILFRLDIRMGLLYITLC 120
|||||
QY 84 -MFLMTCKPPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANMSNDCQSFAPYIADL 142
|||||
Db 121 IYFLMTCKPPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANMSNDCQSFAPYIADL 180
|||||
QY 143 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGGEAMRRPQIDKKGR 202
|||||
Db 181 SLKYNTGLNFGKVDVGRYTDVSTRYKYSTPLTKOLPTLLILFOGGEAMRRPQIDKKGR 240
|||||
QY 203 AVSWTFSEENVIREFNELNELYORAKKLSKAGDNIPPEQVASTPTTVSDGENKKDK 258
|||||
Db 241 AVSWTFSEENVIREFNELNELYORAKKLSKAGDNIPPEQVASTPTTVSDGENKKDK 296
|||||

RESULT 12

US-10-173-706-32

; Sequence 32, Application US/10173706

; Publication No. US20030022293A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3430R1C7

; CURRENT APPLICATION NUMBER: US/10/173,706

; CURRENT FILING DATE: 2002-06-17

; PRIOR APPLICATION removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 32

; LENGTH: 296

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-173-706-32

Query Match 97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

```
Qy 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Db 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNFMFSKVANTILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
```

RESULT 13

```
---10-175-738-32
sequence 32, Application US/10175738
Publication No. US20030022294A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC45
; CURRENT APPLICATION NUMBER: US/10/175,738
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-32
```

Query Match 97.6%; Score 1326; DB 9; Length 296;

Best Local Similarity 86.8%; Pred. No. 1.2e-129;

Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

```
Qy 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Db 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNFMFSKVANTILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
```

RESULT 14

```
US-10-175-752-32
sequence 32, Application US/10175752
Publication No. US20030022295A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC60
; CURRENT APPLICATION NUMBER: US/10/175,752
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 32
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-32
```

Query Match 97.6%; Score 1326; DB 9; Length 296;

Best Local Similarity 86.8%; Pred. No. 1.2e-129;

Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

```
Qy 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Db 1 MAVLAPLIALVSVPRLSRWLAQPYLLSALLSAAFLVRLKLPPLCHGLPTOREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMMKNRRS----- 83
Db 61 FDWREVEILMFLSAIVMMKNRRSITVEQHIGNFMFSKVANTILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELERDKRVTWIVFFANWSNDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSPLTKQLPTLLILFOGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNLNELYORAKKLSKAGDNIPPEQPVASTPTTVSDGENKKDK 296
```

RESULT 15

```
US-10-176-482-32
sequence 32, Application US/10176482
Publication No. US20030022296A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```

: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C70
: CURRENT APPLICATION NUMBER: US/10/176,482
: CURRENT FILING DATE: 2002-06-20
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 32
: LENGTH: 296
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-176-482-32

Query Match      97.6%; Score 1326; DB 9; Length 296;
Best Local Similarity 86.8%; Pred. No. 1.2e-129;
Matches 257; Conservative 1; Mismatches 0; Indels 38; Gaps 1;

      1 MAVLAPLIAIVYSPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
      |||
      1 MAVLAPLIAIVYSPRLSRWLAQPYLLSALLSAAFLLVKRLPPLCHGLPTQREDGNPCD 60
      |||
      61 FDWREVEILMFLSAIVMMKNRRS----- 83
      |||
      61 FDWREVEILMFLSAIVMMKNRRSITVEQHIQNIFFMFSKVANTILFFRLDIRMGLLYITLC 120
      :|||
      84 -MFLMTCCKPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCOSFAPYIADL 142
      |||
      121 IVFLMTCCKPLYMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANWNSDCOSFAPYIADL 180
      |||
      143 SLKYNTGTLNFGKVDVGRYTDVSTRYKVSTSPLTQLPTLILFOGKREAMRRPOIDKKGR 202
      |||
      181 SLKYNTGTLNFGKVDVGRYTDVSTRYKVSTSPLTQLPTLILFOGKREAMRRPOIDKKGR 240
      |||
      203 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPEEQPVASTPTTVSDGENKKDK 258
      |||
      241 AVSWTFSEENVIREFNELNYQRAKLSKAGDNIPEEQPVASTPTTVSDGENKKDK 296

Search completed: July 9, 2003, 11:58:31
Job time : 25 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 12, 2003, 10:39:22 ; Search time 1105 Seconds
(without alignments)
3781.390 Million cell updates/sec

Title: US-09-954-846-2
Perfect score: 1359
Sequence: 1 MAVLAPLIALVYSPRLSRW.....EQPVASTPTTVSDGENKKDK 258

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09954846/runat_09072003_102145_12167/app.query.fasta_1.455
-DB=EST -QWTF=fastap -SUFFIX=rst -HINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.coi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09954846 -CGN_1_1456_runat_09072003_102145_12167 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tabase :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_hic:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oth:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1354	99.6	1022	13	BM555220	BM555220 AGENCOURT
2	1335	98.2	919	14	BQ961752	BQ961752 AGENCOURT
3	1318	97.0	884	14	BQ430136	BQ430136 AGENCOURT
4	1302	95.8	893	9	AL543426	AL543426 AL543426
5	1284	94.5	913	14	BQ920164	BQ920164 AGENCOURT
6	1265	93.1	1020	14	BM923588	BM923588 AGENCOURT
7	1255.5	92.4	793	13	BI771888	BI771888 603055295
8	1254	92.3	907	9	AL530852	AL530852 AL530852
9	1251	92.1	1097	14	BQ073145	BQ073145 AGENCOURT
10	1246.5	91.7	1030	14	BM905350	BM905350 AGENCOURT
11	1243	91.5	875	14	BQ937601	BQ937601 AGENCOURT
12	1241	91.3	1076	14	BM921414	BM921414 AGENCOURT
13	1231	90.6	938	14	BQ717825	BQ717825 AGENCOURT
14	1223	90.0	883	14	BQ439896	BQ439896 AGENCOURT
15	1220.5	89.8	1969	11	AK009759	AK009759 MUS MUSCU
16	1217.5	89.6	859	9	AL544734	AL544734 AL544734
17	1210	89.0	858	14	BQ230217	BQ230217 AGENCOURT
18	1209.5	89.0	890	9	AL515717	AL515717 AL515717
19	1199	88.2	724	12	BE741788	BE741788 601595590
20	1183	87.0	1057	13	BM556247	BM556247 AGENCOURT
21	1179	86.8	948	12	BE796785	BE796785 601587565
22	1173	86.3	1006	13	BM562846	BM562846 AGENCOURT
23	1171	86.2	1053	14	BQ070250	BQ070250 AGENCOURT
24	1170	86.1	933	14	BM974314	BM974314 UT-CF-EC1
25	1159	85.3	767	14	BQ957743	BQ957743 AGENCOURT
26	1159	85.3	942	14	BQ957743	BQ957743 AGENCOURT
27	1154	84.9	991	14	BQ691103	BQ691103 AGENCOURT
28	1148	84.5	969	13	BM450430	BM450430 AGENCOURT
29	1143	84.1	858	14	BQ686696	BQ686696 AGENCOURT
30	1139	83.8	900	13	BI838873	BI838873 603087214
31	1136	83.6	937	13	BI598053	BI598053 603248183
32	1132.5	83.3	883	13	BI335375	BI335375 602979737
33	1126.5	82.9	939	14	BQ938032	BQ938032 AGENCOURT
34	1125	82.8	887	14	BQ887034	BQ887034 AGENCOURT
35	1123	82.8	894	13	BI654405	BI654405 603280668
36	1118	82.3	811	13	BI550022	BI550022 603192682
37	1115	82.0	830	13	BI764082	BI764082 603043295
38	1109	81.6	763	9	AL545332	AL545332 AL545332
39	1108	81.5	797	13	BI521249	BI521249 603081808
40	1108	81.5	905	13	BI836556	BI836556 603089519
41	1091.5	80.3	892	14	BQ430854	BQ430854 AGENCOURT
42	1086	79.9	894	13	BI597723	BI597723 603248377
43	1085.5	79.9	937	13	BI550817	BI550817 603195522
44	1084	79.8	858	12	BG831608	BG831608 602765872
45	1077	79.2	865	13	BI253341	BI253341 602973613

ALIGNMENTS

RESULT 1
BM555220
LOCUS
DEFINITION AGENCOURT_6541254 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5550270
5', mRNA sequence.
ACCESSION BM555220
VERSION BM555220.1 GI:18795504
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

QY 1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTrrp 20
 Db 23 ATGGCGGTCTGGCACCTTAATTGCTCGTGTAATTCGGTGGCGGACATTTCCACGATGG 82
 QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
 Db 83 CTCGCCCAACCTTACTACCTTCTGTGGCCCTGCTCTGCTGCTCTACTCGTGAGG 142
 QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 Db 143 AAATGCGCGCCCTCTGCCACGCTGTGCCACCAACGCAAGACGCTAACCCGCTGTGAC 202
 QY 61 PheAspTrrpArgGluValGlnLeuMetPheLeuSerAlaLeuValMetLysAsn 80
 Db 203 TTGTACTGGAGAGAGTGGAGATCCGATGTTCTCAGTGCCCATGTGTGATGAGAGAAC 262
 QY 81 ArgArgSerMetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIle 100
 Db 263 CGCAGATCCATGTTCTGTGATGACGTGCAAAACCCCTATATATGGCCCTGAGTATATC 322
 QY 101 LysTyrPheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrrp 120
 Db 323 AAGTACTTCAATGATAAAACCATTTGATGAGAACTAGAACGGGACCAAGAGGTCACCTGG 382
 QY 121 IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAla 140
 Db 383 ATGTGGAGATCTTTGGCAATTTGGTCTAATGATGCGCAATCATTTGCCCTCTCTATGCT 442
 QY 141 AspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArg 160
 Db 443 GACCTCTCCCTTAATACAACTGTACAGGCTAAATTTTGGGAAGTGGATGTGGACGC 502
 QY 161 TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro 180
 Db 503 TATACTGATGTTAGTACGGGTACAAAGTGAGCAGATCACCCCTCACCAAGCAATCCCT 562
 QY 181 ThrLeuIleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
 Db 563 ACCCTGATCTGTTCCAGGTGGCAAGGAGCAATGCGGGGCCACAGATTGACAAGAAA 622
 QY 201 GlyArgAlaValSerTrrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn 220
 Db 623 GGACGGGCTGCTCATGACCTTCCTGAGGAGAAATGTGATCGGAGAAATTTAACTTANAT 682
 QY 221 GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGln 240
 Db 683 GAGCTATACCAAGCGGGCCCAAGAACTATCAAAGGCTGGAGACAATATCCCTGANGAGCAG 742
 QY 241 ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAsp 257
 Db 743 CCTGTGGCTTCAACCCCAACCACTGTCAGATGGGGAACNAGAAGGAT 793
 RESULT 3
 BQ430136
 LOCUS
 DEFINITION AGNCOURT_7809582 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6053487
 5', mRNA sequence.
 ACCESSION BQ430136
 VERSION BQ430136.1 GI:21169212
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (Bases 1 to 884)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 plate: LLAM13310 row: j column: 16
 High quality sequence stop: 599...
 FEATURES
 Location/Qualifiers
 1..884
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6053487"
 /clone_lib="NIH_MGC_70"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 BASE COUNT 216 a 235 c 223 g 209 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.74e-155 Length: 884
 Score: 1318.00 Matches: 251
 Percent Similarity: 98.44% Conservative: 2
 Best Local Similarity: 97.67% Mismatches: 4
 Query Match: 96.98% Indels: 0
 DB: 14 Gaps: 0
 US-09-954-846-2 (1-258) x BQ430136 (1-884)

QY 1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTrrp 20
 Db 37 ATGGCGGTCTGGCACCTTAATTGCTCGTGTAATTCGGTGGCGGACATTTCCACGATGG 96
 QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
 Db 97 CTCGCCCAACCTTACTACCTTCTGTGGCCCTGCTCTGCTGCTCTCTACTCGTGAGG 156
 QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 Db 157 AAATGCGCGCCCTCTGCCACGCTGTGCCACCAACGCAAGACGCTAACCCGCTGTGAC 216
 QY 61 PheAspTrrpArgGluValGlnLeuLeuMetPheLeuSerAlaIleValMetLysAsn 80
 Db 217 TTGTACTGGAGAGAACTGGAGATCCCTGATGTTCTCAGTGCCCATTTGTGATGATGAGAAG 276
 QY 81 ArgArgSerMetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIle 100
 Db 277 CGCAGATCCATGTTCTGTGATGACGTGCAAAACCCCTATATATGGCCCTGAGTATATC 336
 QY 101 LysTyrPheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrrp 120
 Db 337 AAGTACTTCAATGATAAAACCATTTGATGAGGAACTAGAACGGGACCAAGAGGTCACCTGG 396
 QY 121 IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAla 140
 Db 397 ATTTGGAGTCTTTTGGCAATTTGGTCTAATGATGACTGCCAATCATTTGCCCTCTATGCT 456
 QY 141 AspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArg 160
 Db 457 GACCTCTCCCTTAATACAACTGTACAGGCTAAATTTTGGGAAGGTTGATGTGGACGC 516
 QY 161 TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro 180
 Db 517 TATACTGATGTTAGTACGGGTACAAAGTGACACATCACCCCTCACCAAGCAATCCCT 576
 QY 181 ThrLeuIleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
 Db 577 ACCCTGATCTCTTCCAAAGTGGCAAGGAGCAATGCGGGCCACAGATTGACAAGAAA 636
 QY 201 GlyArgAlaValSerTrrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn 220

Db 637 GGAGGGCTGCTCATGACCTTCTCTNGAGAGATGTGATCCGAGATTTAACTTAAT 696
 Qy 221 GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGln 240
 Db 697 GAGCTATACGAGGGGCGAAGAACTATCAAGGCTGGAGACATATCCCTGAGGAGCAG 756
 Qy 241 ProValAlaSerThrProThrValSerAspGlyGluAsnLysLysAsp 257
 Db 757 CTTGGCTTCAACCCACCACCGCTGTCAGATGGGGAACAAGAGGGAT 807

RESULT 4
 AL543426 893 bp mRNA linear EST 16-FEB-2001
 LOCUS AL543426 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1003YH21 5
 DEFINITION prime, mRNA sequence.

ACCESSION AL543426
 VERSION AL543426.1 GI:12875904
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 893)

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers

1..893

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0D1003YH21"

/clone_lib="LTI_NFL006.PL2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 233 a 228 c 206 g 224 t 2 others
 ORIGIN

Annotation Scores:

i. No.: 4.99e-153 Length: 893
 Matches: 252
 Percent Similarity: 86.64% Conservative: 1
 Best Local Similarity: 86.30% Mismatches: 1
 Query Match: 95.81% Indels: 38
 DB: 9 Gaps: 1

US-09-954-846-2 (1-258) x AL543426 (1-893)

Qy 5 AlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTyrLeuAlaGlnPro 24
 Db 3 GCACCTCTAATGCTCTCGTGATTCGGTGGCGGACTTTCACCATGGCTCGCCACCT 62
 Qy 25 TyrTyrLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArgLysLeuPro 44
 Db 63 TACTACCTTCTGTGGCCCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122
 Qy 45 LeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAspPheAspTyrArg 64
 Db 123 CTCTGCCACGGTCTGCCACCCACCGGAACGCGTAACCGGTGCTGTGACTTGGAGA 182

Qy 65 GluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsnArgSer --- 83
 Db 183 GAATGGAGATCCCTGATGTTTCTCAGTCCCATGTGATGATGAAGAACCGAGATCCATC 242
 Qy 83 ----- 83
 Db 243 ACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAAAGTGGCCCAACAACATTTCT 302
 Qy 84 -----MetPheLeu 86
 Db 303 TCTTCCCTTGGATATTTCGCATGGCCCTACTTTACATCACACTCTGCATAGTGTCTCTG 362
 Qy 87 MetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyrPheAsnAspLys 106
 Db 363 ATGACGTCAAAACCCCTATATATGGCCCTGAGTATATCAAGTACTTCAATGATAAA 422
 Qy 107 ThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleValGluPheAla 126
 Db 423 ACCATTGATAGGAACACTAGAACGGGACAGAGGGTCACTTGGATTGTGGAGTCTTTGCC 482
 Qy 127 AsnTyrSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLysTyr 146
 Db 483 AATTGGTCTATGACTGCAATCATTTGCCCTATCTATGCTGACTCTCCCTTAAATAC 542
 Qy 147 AsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThrAspValSerThr 166
 Db 543 AACTGTACAGGGCTAAATTTGGGAAGTGGATGTTGGACGCTATCTATCTAGTACG 602
 Qy 167 ArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGln 186
 Db 603 CGGTACAAGTGAGCACATCACTCCCTCACCAGCACTCCCTACCTGATCTCTGTCCAA 662
 Qy 187 GlyGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSerTyr 206
 Db 663 GGTGCAAGGAGGCAATCGCGGCCACAGATTGACAGAAAGGACGGCTCTCTATGG 722
 Qy 207 ThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAla 226
 Db 723 ACCTTCTCTGAGGAGAATGTGATCGAGAAATTTAACTTAATGAGCTATACCGAGGCC 782
 Qy 227 LysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThrPro 246
 Db 783 AAGAACTATCAAGGGCTGGAGAAATATCCCTGAGGACAGCCCTGTGGCTTCAACCCCC 842
 Qy 247 ThrThrValSerAspGlyGluAsnLysLysAspLys 258
 Db 843 ACCACAGTCTCAGTGGGAAACAGAGGATATAA 878

BO920164 913 bp mRNA linear EST 20-AUG-2002
 AGENCOURT_8918912 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6383675
 5', mRNA sequence.

ACCESSION BO920164

VERSION BO920164.1 GI:22334862

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 913)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov


```

Db      437 TGACCTCTCCCTTAATACAACTGTACAGGCTAAATTTTGGGAAGGTGGATGTTGGAGC 496
Qy      160 gTyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPr 180
Db      497 CTATACCTGATGTTAGTACGGGTCAAAAGTACAGACATCACCCCTCCACCAAGCAACTCCC 556
Qy      180 oThrLeuLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys 200
Db      557 TACCCTGATCTGTTCCTCAAGGTGGCAAGAGGCAATGCGGGGCA-CAGATTGACAAGAA 615
Qy      200 sGlyArgAlaValSerThrPheSerGluGluAsnValIleArgGluPheAsnLeuAs 220
Db      616 AGGAGGGCTGCTCATGAGCTTCTCTGAGAGAGATGTGATCCGAGAAATTAACCTTAAA 675
Qy      220 nGluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGlu 240
Db      676 TGAGCTATACAGCGGGC-AAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCA 734
Qy      240 nProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAsp 257
Db      735 GCCTGTGCTTCAACCCAC--ACAGTGTCTAGATGGGGAACAAGAGGAT 784

RESULT 8
AL530852 907 bp mRNA linear EST 13-FEB-2001
LOCUS     AL530852 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD008YG11 5
DEFINITION prime, mRNA sequence.
ACCESSION AL530852
VERSION   AL530852.1 GI:12794345
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 907)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
           source
           1..907
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="CS0DD008YG11"
           /clone_lib="LTI_NFL001_NBC4"
           /sex="male"
           /tissue_type="neuroblastoma cells"
           /lab_host="DH10B"
           /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
           was primed with a NotI-oligo(dT) primer. Five prime end
           enriched, double-stranded cDNA was digested with Not I and
           cloned into the Not I and Eco RV sites of the pCMVSPORT 6
           vector. Library was normalized. Library was constructed
           by Life Technologies. Contact : Feng Liang Life
           Technologies, a division of Invitrogen 9800 Medical Center
           Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
           8371 Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com"

BASE COUNT 232 a 231 c 208 g 228 t 8 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5.66e-147      Length:      907
Score:          1254.00      Matches:      245
Percent Similarity: 84.48%      Conservative: 0
Best Local Similarity: 84.48%      Mismatches: 7
Query Match:    92.27%      Indels:      38
DB:              9      Gaps:      2

```

```

US-09-954-846-2 (1-258) x AL530852 (1-907)
Qy      7 LeuIleAlaLeuValTyrSerValProArgLeuSerArgTyrTrpLeuAlaGlnProTyrTyr 26
Db      3 CTAATTCGTCTCGTGTATTCGGTCCGGGACTTTCCAGTATGCTCGCCCAACTTACTACT 62
Qy      27 LeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArgLysLeuProLeuCys 46
Db      63 CTTCTGTCTGGCCCTCTCTCTGCTGCTTCTTACTCTGTGAGGNAAACTGCCCGCGCTCTGC 122
Qy      47 HisGlyLeuProThrGlnArgGluAspGlyAsnProCysAspPheAspThrPargGluVal 66
Db      123 CACGCTCTGCCCAACCAACGCGAAGACGTAACCCGTGTGACTWTGACTGGAGAGAGTG 182
Qy      67 GluIleLeuMetPheLeuSerAlaIleValMetMetLysAsnArgSer----- 83
Db      183 GAGATCCTCGATGTTTCTCAGTGCCATTTGTGATGATGATGAAGAACCGAGATCCCATCTG 242
Qy      84 -----MetPhe----- 85
Db      243 GAGCAACATATAGGCAACATTTTTCATGTTTAAAGTGGCCCAACACAAATTTCTTCTTC 302
Qy      86 -----LeuMetThr 88
Db      303 CCCTTGGATATTCGCATGGCCCTACTTTACATCACACTCTGCATAGCTGTWCCNGAAGCG 362
Qy      89 CysLysProProLeuTyrMetGlyProGluTyrIleLysTyrPheAsnAspLysThrIle 108
Db      363 TCGAAACCCCCCTATATATGGCCCTGAGTATATCAAGTACTTCAATGATATAAACCAATT 422
Qy      109 AspGluGluLeuGluArgAspLysArgValThrTrpIleValGluPhePheAlaSerTrp 128
Db      423 CATGAGGAACATAGAAAGGAGGTCACCTTGGATTGTGGAGTTCTTTGCCCAATTGG 482
Qy      129 SerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLysTyrAsnCys 148
Db      483 TCTAATGACTGCCAATCATTTGCCCTATCTATGCTGACCTCTCCCTTAAATACAACTGT 542
Qy      149 ThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThrAspValSerThrArgTyr 168
Db      543 ACAGGGCTAAATTTTGGGAAGGTGGATGTTGGACGCTATACTGATGTAGTACGCGGTAC 602
Qy      169 LysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGlnGly 188
Db      603 AAAGTGACGCMCATMACCCCTCACCAAGCAACTCCCTACCTCATCTGTTCCAAGGTGC 662
Qy      189 LysGluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSerThrPhe 208
Db      663 AAGGAGGCAATGCGGCGCCACAGATTGACAAGAAGGACGGGCTGTCTCATGGACCTTC 722
Qy      209 SerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAlaLysLys 228
Db      723 TCTGAGGAGAATGTTATCCGAGAATTTAACTTAAATGAGCYATACCAGCGGCCCAAGAAA 782
Qy      229 LeuSerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThrProThr 248
Db      783 CTATCAAGGCTGGAGACAATATCCCTGAGGAGACGCTGTGGCTTCAACCCCAACCA 842
Qy      249 ValSerAspGlyGluAsnLysLysAspLys 258
Db      843 GTGTCTAGATGGGGAACAAGAGGATAAA 872

RESULT 9
LOCUS     B0073145
DEFINITION AGENCOURT_6817826 NIH_MCC_118 Homo sapiens cDNA clone IMAGE:5757493
           .5', mRNA sequence.
ACCESSION B0073145
VERSION   B0073145.1 GI:19902191
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1097)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM12799 row: m column: 14
 High quality sequence stop: 637.
 Location/Qualifiers
 1..1097
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_118"
 /tissue_type="leukocyte"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
 (destroyed); RNA source leukocytes from anonymous pool of
 non-activated adult donors. Library is oligo-dT primed
 and directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 1.2-3.3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 027. Note:
 this is a NIH_MGC Library."
 BASE COUNT 261 a 296 c 276 g 263 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.8e-146 Length: 1097
 Score: 1251.00 Matches: 246
 Percent Similarity: 97.63% Conservative: 1
 Best Local Similarity: 97.23% Mismatches: 3
 Query Match: 92.05% Indels: 3
 Gaps: 0
 14
 US-09-954-846-2 (1-258) x B0073145 (1-1097)
 QY 3 ValLeuAlaProLeuIleAlaLeuValTySerValProArgLeuSerArgTirPleuAla 22
 DB 1 GTCTTGGCACCTCTAATGCTCTCGTCTCTCGTGGCGGACCTTCACGATGGCTGCC 60
 QY 23 GluProTyrTyrLeuLeuSerAlaLeuSerAlaAlaPheLeuLeuValArgLysLeu 42
 61 CAACCTTACTACCTTCTGTCGGCCCTGCTCTGCTGCTTCCCTTCCCTGAGGAACTG 120
 QY 43 ProProLeuCysHisGlyLeuProThrGluArgGluAspGlyAsnProCysAspPheAsp 62
 DB 121 CCGCCGCTGTCGCCACGCTGCCCCACCCACCGGACCGGTAACCCCTGTGACTTTGAC 180
 QY 63 TirArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsnArgArg 82
 DB 181 TGAGAGAACTGGAGATCTTGATGTTCTCAGTGCCATTGTGATGATGAGAACCGGAGA 240
 QY 83 SerMetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
 DB 241 TCCATGTTCTGTAGCTGCAAAACCCCTATATATGCGCCCTGATGATATCAAGCAC 300
 QY 103 PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrTirPileVal 122
 DB 301 TTCAATGATAAAACCATTTGATGAGGAACCTAGAACGGGACAGAGGGTCACTTGGATTGTG 360
 QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
 DB 361 GAGTCTTTCCTCAATTTGGTCTAATGACTGGCAATCATTTGGCCCTATCTATGCTGACCTC 420

QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
 DB 421 TCCCTTAATACAACTGTACAGGCTAAATTTTGGGAAGTGGATGTGGACGCTATACT 480
 QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
 DB 481 GATGTTAGTACCGGTACAAAGTACAGACATCACCCCTCACCAAGCAACTCCCTACCTG 540
 QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
 DB 541 ATCTGTTCCTCAAGTGGCAAGGAGCAATCGCGGCCACAGATTGACAAGAAGACGG 600
 QY 203 AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
 DB 601 GCTGCTCATGGACCTTCTCTGAGGAGAAATGTGATCCGAGAAATTTAACTTAAATGAGCTA 660
 QY 223 TyrGlnArgAlaLysLysLeuSerLysAla-GlyAspAsnIle-ProGluGluGlnProV 242
 DB 661 TACCCGGGGCCCAAGAACTATCAAGGCTGGGAGACAATATCTCTGAGGAGACACCTG 720
 QY 242 alAla-SerThrProThrThrValSerAspGly 252
 DB 721 TGGCCTTCAACCCCCACCACAGTGTGACATGGG 753
 RESULT 10
 BM905350
 LOCUS AGENCOURT_6698724 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5558249
 DEFINITION 5', mRNA sequence.
 ACCESSION BM905350
 VERSION BM905350.1 GI:19355729
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1030)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM12280 row: o column: 18
 High quality sequence stop: 711.
 Location/Qualifiers
 1..1030
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 254 a 274 c 245 g 255 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.03e-146 Length: 1030
 Score: 1246.50 Matches: 249
 Percent Similarity: 84.90% Conservative: 4
 Best Local Similarity: 83.56% Mismatches: 4
 Query Match: 91.72% Indels: 41


```

DB:      14      Gaps:      2
US-09-954-846-2 (1-258) x BM905350 (1-1030)

QY      1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTrrp 20
DB      36 ATGGGGGCTTGGCACTTAATATGCTCGTGTATTCGGTGGCGGACTTTCCACGATGG 95
QY      21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
DB      96 CTCGCCCAACCTTACTACTCTTCGTGGCCCTGCTCTGCTGCCCTTCTACTCGTGAGG 155
QY      41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
DB      156 AAATGCGCCGCTCTGCGCACGCTCTGCCACCGCTCTGCCACCGCAAGACGCGTAAACCGGTGAC 215
QY      61 PheAspTrrArgGluValGluLeuLeuMetPheLeuSerAlaAlaValMetMetLysAsn 80
DB      216 TTGTGACTGGAGAGAGTGAGATCCTGTATGTTTCTCAGTGCCATTGTGTATGATGAAGAAC 275
QY      81 ArgArgSer----- 83
DB      276 CGCAGATCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGGCC 335
QY      83 ----- 83
DB      336 AACACAATCTTTCTTCGCTTGGATATTCGCATGGGCGCTACTTTACATCACACTCTGC 395
QY      84 ----MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
DB      396 ATAGTGTTCCTGATGACGTGCAAAACCCGCCCTATATATGGGCCCTCAGTATATCAAGTAC 455
QY      103 PheAsnAspLysThrIleAspGluLeuLeuGluLeuArgAspLysArgValThrTrpIleVal 122
DB      456 TTCAATGATATAAACCACTTATGAGGAACATAGAGCAAGGAGGCTCACTTGGATGTGG 515
QY      123 GluPhePheAlaAsnTrrSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
DB      516 GAGTCTTTGCCAATTTGGTCTAATGACTGCCAATCATTTGCCCTATCTATGCTGACCTC 575
QY      143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
DB      576 TCCCTTAATAACAACGTACAGGGCTAAATTTTGGGAAGTGGATCTTGGACGCTATACT 635
QY      163 AspValSerThrArgTyrIleValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
DB      636 GATGTAGTACCGGTACAAAGTGAGCACATCACCCCTCACCAGCAACTCCCTACCCCTG 695
QY      183 IleLeuPheGlnGlyGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
DB      696 ATCTGTTCGAAGGTGGCAAGGAGCAATGCGGGCCACAGATTGACAAAGAACGACGG 755
QY      203 AlaValSerTrrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
DB      756 GCTGTCTCATGGACCTTCTCTCAAGAGATGTGATCCGAGATTAACTTAAATGAGCTA 815
QY      223 TyrGlnArgAlaLysLysLeuSerLysAla-GlyAspAsnIleProGlu-GluGlnProV 242
DB      816 TACCAGCGGGCCCAAGAACTATCAAGGCTGGGAGACAAATATCCCCCTGAGAGCAGCCTG 875
QY      242 AlaIleSerThrProThrThrValSerAsp--GlyGluAsnLysLysAsp 257
DB      876 TGGCTTTTACCCCCACACAGTGGGCGAGATGGGGGAAAAACAGNAGGGAT 925

RESULT 11
LOCUS   BQ937601
DEFINITION AGENCOURT_8925729 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6470473
5', mRNA sequence.
ACCESSION BQ937601
VERSION   BQ937601.1 GI:22352972
KEYWORDS EST.
SOURCE   human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
  1 (bases 1 to 875)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabsb@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  Cloning Strategy: Agencourt Bioscience Corporation
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM14001 row: e column: 02
  High quality sequence stop: 684.
  Location/Qualifiers
    1..875
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone_lib="IMAGE:6470473"
     /clone_lib="NIH_MGC_71"
     /tissue_type="leiomyosarcoma"
     /lab_host="DH10B (phage-resistant)"
     /note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
     Average insert size 2.1 kb."
  BASE COUNT 219 a 229 c 208 g 219 t
  ORIGIN

Alignment Scores:
  Pred. No.:      1,3e-145      Length:      875
  Score:          1243.00      Matches:    244
  Percent Similarity: 85.42%      Conservative: 2
  Best Local Similarity: 84.72%      Mismatches: 3
  Query Match:     91.46%      Indels:    39
  DB:              14          Gaps:      1

US-09-954-846-2 (1-258) x BQ937601 (1-875)

QY      1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTrrp 20
DB      12 ATGGCGGCTTGGCACTTAATATGCTCGTGTATTCGGTGGCGGACTTTCCACGATGG 71
QY      21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg 40
DB      72 CTCGCCCAACCTTACTACTCTTCGTGGCCCTCTCTCTGCTGCCCTTCTACTCGTGAGG 131
QY      41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
DB      132 AAATGCGCGGCTCTGCCACGCTCTGCCACCGCAAGACGCGTAAACCGGTGAC 191
QY      61 PheAspTrrArgGluValGluLeuLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
DB      192 TTGTGACTGGAGAGAGTGAGATCCTGTATGTTTCTCAGTGCCATTGTGTATGAAGAAC 251
QY      81 ArgArgSer----- 83
DB      252 CGCAGATCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGGCC 311
QY      83 ----- 83
DB      312 AACACAATCTTTCTTCGCTTGGATATTCGCATGGGCGCTACTTTTACATCACACTCTGC 371
QY      84 ----MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
DB      372 ATAGTGTTCCTGATGACGTGCAAAACCCGCCCTATATATGGGCCCTGAGTATATCAAGTAC 431
QY      103 PheAsnAspLysThrIleAspGluLeuLeuGluLeuArgAspLysArgValThrTrpIleVal 122
DB      432 TTCATGATAAAACCACTTATGAGGAACATAGACGGGACAGAGGGGTCACTTGGATGTGTG 491

```

```

QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db 492 GAGTTCCTTTGCCAATTGGTCTAATGACTGCGCAATCATTTGCCCTATCTATGCTGACCTC 551
QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPhcGlyLysValAspValGlyArgTyrThr 162
Db 552 TCCCTTAATACAACTGTACAGGCTAAATTTGGGAAGGTGGATGTTGGACGCTACTACT 611
QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db 612 GATGTTAGTACGGGTACAAAGTGAGCAGACATCACCCTCACCAGCAACTCCCTACCTG 671
QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db 672 ATCCTGTTCAGGTGGCAAGGAGGCAATCGCGGCCACAGATGTGACAAAGAAAGGACGG 731
QY 203 AlaValSerTrpThrPheSerGluAlaValIleArgGluPheAsnLeuAsnGluLeu 222
Db 732 GCTGTCTCATGGACCTTCTCTGAGGAGATGTGATCCGAGAAATTAATTTAATGAGCTA 791
QY 223 TyrGlnArgAlaLys-LysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVa 242
Db 792 TACCAGCGGCCCAAGAAACTATCAAGGCTGGAGACAATATCCTGAGGAGCAGCCTGT 851
QY 242 lAlaSerThrProThrThrVal 249
Db 852 GGCTTCAACCCACACAGGTG 873

RESULT 12
LOCUS BM921414 1076 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6626480 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5752847
5', mRNA sequence.
ACCESSION BM921414
VERSION BM921414.1 GI:19371793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1..(bases 1 to 1076)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M12787 row: k column: 24
High quality sequence stop: 697.
Location/Qualifiers
1..1076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5752847"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dr primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."
FEATURES
source

```

```

BASE COUNT 257 a 296 c 251 g 270 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 3.17e-145 Length: 1076
Score: 1241.00 Matches: 253
Percent Similarity: 85.28% Conservative: 2
Best Local Similarity: 84.82% Mismatches: 1
Query Match: 91.32% Indels: 43
DB: 14 Gaps: 1
US-09-954-846-2 (1-258) x BM921414 (1-1076)
QY 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp 20
Db 18 ATGCGCGTCTTGGACCTCTAATGCTCTCGTATTGCGTGGCGCGACTTTCACGATGG 77
QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
Db 78 CTCGCCCAACCTTACTACCTTCTGCGGCCCTGCTCTGCTGCTTCTTCTACTCGTGAGG 137
QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
Db 138 AAACCTGCGCGCTCTGCGACGGTCTGCCACCGCTGCCACCGCGGAAGACGGTAACCCGTGTGAC 197
QY 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
Db 198 TTTGACTGGAGAGAAGTGGAGATCTGATGTTTCTCAGTGCCTTGTGATGAAGAAC 257
QY 81 ArgArgSer----- 83
Db 258 CGCAGATCCATCACTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGGCC 317
QY 83 ----- 83
Db 318 AACACAATTCCTTTCTCCGCTTGGATATTCGCATGGCGCTACTTTTACATCACACTCTGC 377
QY 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
Db 378 ATAGTGTTCCTGTGATGACGTGCAACCCCTCTATATATGCGCCCTGAGTATATCAAGTAC 437
QY 103 PheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTripleVal 122
Db 438 TTCATATATAAACCACTGATGAGAACTAGACGGGACAGAGGGTGCCTTGGATGTG 497
QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db 498 GAGTTCCTTTGCCAATTTGGTCTAATGACTGCGCAATCATTTGCCCTATCTATGCTGACCTC 557
QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPhcGlyLysValAspValGlyArgTyrThr 162
Db 558 TCCCTTAATACAACTGTACAGGCTAAATTTGGGAAGGTGGATGTTGGACGCTACTACT 617
QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db 618 GATGTTAGTACCGGTACAAAGTGAGCAGACATCACCCTCACCAGCAACTCCCTACCTG 677
QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db 678 ATCCTGTTCAGGTGGCAAGGAGCAATCGCGGCCACAGATTTGACAAAGAAAGGACGG 737
QY 203 AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
Db 738 GCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCCGAGAAATTAATTAATGAGCTA 797
QY 223 TyrGlnArgAla-LysLysLeuSer-LysAlaGlyAspAsnIle-ProGluGluGlnPro 241
Db 798 TACCAGCGGCCCAAGAAACTATCAAGGCTGGAGACAATATTTCCCTTGAGGAGCAGCCT 857
QY 242 Val-AlaSerThrProThrThrValSerAsp-GlyGluAsnLysLys 256
Db 858 GTGGCTTCAACCCACCGACAGTGGCGGGAACACAGAG 904

```

```

RESULT 13
B0717825
LOCUS      B0717825      938 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION AGENCOURT_8229576 Lupski_dorsal_root_ganglion Homo sapiens cDNA
            clone IMAGE:6184859 5', mRNA sequence.
ACCESSION  B0717825
VERSION     B0717825.1 GI:21856722
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 938)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
IDENT       Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13574 row: 1 column: 12
            High quality sequence stop: 684.

FEATURES             source
     source           1..938
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6184859"
                     /clone_lib="Lupski_dorsal_root_ganglion"
                     /sex="male"
                     /tissue_type="dorsal root ganglia"
                     /dev_stage="adult, 36 yr"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
NOT: Site_2: SalI; cDNA made by oligo-dr priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCG-3' and
5'-GACTAGTCTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."

BASE COUNT  238 a 244 c 221 g 235 t
ORIGIN
Alignment Scores:
Pred. No.:      4.69e-144      Length:      938
Score:          1231.00        Matches:    250
Percent Similarity: 84.28%      Conservative: 2
Best Local Similarity: 83.61%   Mismatches: 4
Query Match:    90.58%         Indels:    43
DB:             14            Gaps:      1

US-09-954-846-2 (1-258) x B0717825 (1-938)

QY      1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp 20
Db      8 ATGGCGGTCTGGCACCTTAATGCTCTGTTATTCGGTGGCGGCACTTTCACGATGG 67
QY      21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
Db      68 CTGCGCCAACTTACTACCTTCTGTCGCGCCCTGCTCTGCTGCTTCTCTACTCGTGAGG 127
QY      41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
Db      128 AAACGCGCGCGCTCTGCCACGGTCTGCCACCAACCGCAAGCGGTAAACCCGTGTGAC 187
QY      61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80

```

```

Db      188 TTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCCATGTGTGATGAAGAAC 247
QY      81 ArgArgSer-----
Db      248 CGCAGATCCATCACTGCTGGGAGCAACATATAGGCAACATTTTTCATGTTTAGTAAAGTGCC 307
QY      83 -----
Db      308 AACCAAAATCTTTTCTCCGCTTGATATTCCGATGGGCTACTTTACATCACACTCTGC 367
QY      84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
Db      368 ATAGTGTCTTCGATGACGTGCAAAACCCCTATATATATGGGCCCTGAGTATATCAAGTAC 427
QY      103 PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrTrpIleVal 122
Db      428 TTCAATGATAAAACCATTTGATGAGCAACTAGAACGGGCAACAGAGGCTCACTTGGATTGG 487
QY      123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db      488 GAGTTCTTTGGCAATTTGCTTAATGACTGCCAATCATTTGCCCTTATCTATGCTGACCTC 547
QY      143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
Db      548 TCCCTTAATACAACTGTACAGGGCTAAATTTGGGAAGGTGGATGTTGGACGCTATACT 607
QY      163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db      608 GATGTTAGTACGGGTACAAAGTAGGACACATCACCCCTCACCAAGCAACTCCCTACCCCTG 667
QY      193 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
Db      668 ATCTGTTTCCAAAGTGGCAAGGAGCAATGCGCGCGCCACAGATTGCACAAGAAAGACGG 727
QY      203 AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
Db      728 CTGCTCTCATGGACCTTCTCTGAGGAGAGATGTGATCCGAGAAATTTAACTTAATAGCTA 787
QY      223 TyrGlnArg-AlaLysLysLeuSerLys-AlaGlyAspAsnIlePro--GluGluGlnPr 241
Db      788 TACCGCGCGGCCCAAGAAACTATCAAGGGCTGGAGACAATATTCCCTGGAGAGACGCC 847
QY      241 oValAlaSerThrProThrVal-SerAspGlyGluAsnLysLys 256
Db      848 TGTGCTTCAAAACCCCAACACAGTTGTCCGATGGGGAACACGAAA 894

RESULT 14
B0439896
LOCUS      B0439896      893 bp      mRNA      linear      EST 24-MAY-2002
DEFINITION AGENCOURT_7809050 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6069357
            5', mRNA sequence.
ACCESSION  B0439896
VERSION     B0439896.1 GI:21178972
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 893)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13351 row: 0 column: 22

```

FEATURES	High quality sequence stop: 616.
source	Location/Qualifiers
	1. .883
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:6069357"
	/clone_lib="NIH_MGC_92"
	/tissue_type="embryonal carcinoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
BASE COUNT	223 a 2230 c 210 g 219 t 1 others

Alignment Scores:	4.37e-143	Length:	893
Pred. No.:	1223.00	Matches:	245
Score:	84.83%	Conservative:	1
Percent Similarity:	84.83%	Mismatches:	3
Best Local Similarity:	84.48%	Indels:	41
Query Match:	89.99%	Gaps:	1
	14		

US-09-954-846-2 (1-258) x BQ439896 (1-883)

Qy	4	LeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrpLeuAlaGln	23
Db	2	TTGGCACCCTCTAAATTGGCTCTCGTATTTCGGTGGCGGACTTTTCACGATGGCTGCCCAA	61
Qy	24	ProTyrTrpLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArgLysLeuPro	43
Db	62	CGTTACTACCTTCTGTGGCCCTGCTCTCTGCTGCTTCTACTCGTAGGAAACTGGCG	121
Qy	44	ProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAspPheAspTrp	63
Db	122	CCGCTCTCCACGGTCTGCCCAACCCACGACGAGCGTACCCGCTGTGACTTTGACTGG	181
Qy	64	ArgGluValGluIleLeuMetPheLeuSerAlaIleValMetLysAsnArgArgSer	83
Db	182	AGAGAAGTGGAGATCCTGATGTTTCTCAGTGCCATTGTGATGATGAAGAACCACGACATCC	241
Qy	83	-----	83
Db	242	ATCACTGTGGAGCAACATATAGGCAACATTTTTCATGTTTAGTAAAGTGCCCAACACAATT	301
Qy	84	-----MetPhe	85
Db	302	CTTTTCTTCGCTTGGATATTCGCATGGCGCTACTTTTACATCACACTCTGCATATAGTTTC	361
Qy	86	LeuMetThrCysLysProLeuTyrMetGlyProGluTyrIleLysTyrPheAsnAsp	105
Qy	362	CTGATGACGTGCAAAACCCCTATATATGGCCCTGAGTATATCAAGTACTTCAATGAT	421
Qy	106	LysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleValGluPhePhe	125
Db	422	AAAACCATTCATGAGGAACTAGACGGGACAGAGGGTCACCTGGATTGTGGAGTTCITT	481
Qy	126	AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys	145
Db	482	GCCAAATTTGGTCTAATGATGCGCAATCATTTTCCGCCCTATCTATGCTGACCTCTCCCTTAAA	541
Qy	146	TyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThrAspValSer	165
Db	542	TACAACGTACAGGGCTAAATTTTGGGAAGTGGATGTTGGACGCTATCTACTGATGTTAGT	601
Qy	166	ThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPhe	185
Db	602	ACGGGTACAAAGTGACCATCACCCCTCACCAAGCAACTCCCTACCTGATCCTGTC	661
Qy	186	GlnGlyGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSer	205

Db	662	CAAGGTGGCAGGAGCAATGCGGGCCACAGATTGACAAGAAAGACGGGCTGTCTCA	721
Qy	206	TpThr-PheSerGluGluAsnValIleAeAGluPheAsnLeuAsnGluLeuTyGlnArg	225
Db	722	TGACCTTCTCTCAAGAGATGTGATCCGAGATTAACTTAAATGAGCTATACACGCG	781
Qy	226	-AlaLysLysLeuSerLysAla-GlyAspAsnIleProGluGluGlnProVal-AlaSer	244
Db	782	GGCCAAGAAACTATCAAGAGCTGGGAGACAATATCCCTCGAGACGACCTCTGGGCTTCA	841
Qy	245	ThrProThrThrValSerAspGly	252
Db	842	ACCCCAACACGATCCGATGG	865
RESULT 15			
LOCUS	AK009759	1969 bp	linear
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310042M24:related to SIMILARITY TO AA 31-55 OF A. THALIANA THIOREDOXINE, full insert sequence.		HTC 19-JAN-2002
ACCESSION	AK009759		
VERSION	1	GI:12844749	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:c57BL/6J) adult male tongue cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Areakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gofjohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikola, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carinci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashina, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H.,

Qy 243 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
Db 847 GACCCCTGCTCCCACTACTGTGTCAGATGGGGAACACAGAGGACAAA 894

Search completed: July 12, 2003, 12:03:35
Job time : 1112 secs


```

Db 61 FDWREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKKDK 296

RESULT 2
Q9H3L1 PRELIMINARY; PRT; 372 AA.
AC Q9H3L1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
% My009 protein.
% Homo sapiens (Human).
% Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Mao Y.M., Xie Y., Zhou Z.X., Ying K., Zheng Z.H.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF059753; AAG43122.1;
DR InterPro; IPR001005; Myb_DNA_Binding.
DR InterPro; IPR000063; ThioRed.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 372 AA; 42477 MW; DE94F9E11B4D8C4C CRC64;

Query Match 92.9%; Score 1263; DB 4; Length 372;
Best Local Similarity 83.3%; Pred. NO. 7.2e-115;
Matches 245; Conservative 4; Mismatches 7; Indels 38; Gaps 1;

Qy 1 MAVLAPLIALVYSPRLSRWLQAOPYLLSALLSAAFLVRLKLPPLCHGLPTQREDGNPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLQAOPYLLSALLSAAFLVRLKLPPLCHGLPTQREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 83
Db 61 FDWREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANTILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKK 256
Db 241 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKK 294

RESULT 3
Q9D710 PRELIMINARY; PRT; 295 AA.
AC Q9D710;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE 2310042M24Rik protein (RIKEN cDNA 2310042M24 gene).
GN 2310042M24Rik.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saifu T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009759; BAB26483.1;
DR EMBL; BC013544; AAH13544.1;
DR MGI; MGI:1914208; 2310042M24Rik.
DR InterPro; IPR001005; Myb_DNA_Binding.
DR InterPro; IPR000063; ThioRed.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 295 AA; 33942 MW; 25E4477F7BB17E3C CRC64;

Query Match 89.8%; Score 1220.5; DB 11; Length 295;
Best Local Similarity 80.4%; Pred. No. 7.4e-111;
Matches 238; Conservative 9; Mismatches 10; Indels 39; Gaps 2;

Qy 1 MAVLAPLIALVYSPRLSRWLQAOPYLLSALLSAAFLVRLKLPPLCHGLPTQREDGNPCD 60
Db 1 MAVLAPLIALVYSPRLSRWLQAOPYLLSALLSAAFLVRLKLPPLCHGLPTQREDGNPCD 60
Qy 61 FDWREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANAILFFRLDIRMGLLYITLC 83
Db 61 FDWREVEILMFLSAIVMKNNRSITVEQHIGNIFMFSKVANAILFFRLDIRMGLLYITLC 120
Qy 84 -MFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 142
Db 121 IVFLMTCKPPLYMGPEYIKYFNNDKTIIDELEDRKRVTVIVEFFANWSDCQSFAPYADL 180
Qy 143 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 202
Db 181 SLKYNCTGLNFGKVDVGRYTVSTRYKYSTSPLTKQLPTLLILFQGGKEAMRRPQIDKKGR 240
Qy 203 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKKDK 258
Db 241 AVSWTFSEENVIREFNELNELYQRAKKLSKAGDNIPEEQVPASTPTTVSDGENKKDK 295

RESULT 4
Q9V8W7 PRELIMINARY; PRT; 271 AA.
AC Q9V8W7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

```

CG11007 protein.

GN Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 5

Q18484

ID Q18484 PRELIMINARY; PRT; 265 AA.

AC Q18484;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Hypothetical 30.9 kDa protein.

GN C35D10.10.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RC STRAIN=BRISTOL N2;

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.;"

RL Science 282:2012-2018(1998).

RN [2]

RC STRAIN=BRISTOL N2;

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "The sequence of C. elegans cosmid C35D10.10.;"

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RC STRAIN=BRISTOL N2;

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Direct Submission.;"

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

RX EMBL: U21324; AAA62566.1; -

DR InterPro; IPR000063; Thioered.

KW Hypothetical protein.

SQ SEQUENCE 265 AA; 30900 MW; 2D0731C464EA5060 CRC64;

Query Match 28.8%; Score 391; DB 5; Length 265;

Best Local Similarity 32.2%; Pred. No. 6e-30;

Matches 84; Conservative 52; Mismatches 81; Indels 44; Gaps

QY 14 VPRLS--RWLAQPYLLSAAFLVLRKLPPLCHGLPTQREDGNPCDFDREVEILMF 71

DB 3 IPRLEDEVRRLATAFHFFNTLLALAPPVIRS-TSLCDYV-FAVEGNEQCEIDSRERILMF 60

QY 72 LSAIVMMKNRRS-----MFLMTC-----KPPL 93

DB 61 LLILAMKGRKATNMHYNNIFLFSKIAGMFLIRADILPGIILACILVTLVLPPEY 120

QY 94 YMGPEYIKYFNDKTIDEELERDKRVTWIVEFFANNSDCQSFAPYADLSLKNCTGLNF 153

DB 121 YNGPEQVITYFOGEQLFEELFRNRNTIWIQFPTWSPCHRTSVFPAELSKQFTLPNNKF 180

QY 154 GKVDGRTYDSTRYKYSTPLTKQPLTLILFOGKGEAMRRPQIDKKGRAVSWTFSENV 213

DB 181 GKLDIGRWAKEGERFRVNAHPSQLFTICVFKDAKEIARRPLVNDSSRRVFPVSENC 240

QY 214 IREFNLNELY--QRAKLKSKA 232

DB 241 VLAFDLLNLYNEQKEKKGAKA 261

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q9LYG6;

DE Drosophila melanogaster (Fruit fly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7221;

RESULT 6

Q9LYG6

ID Q9LYG6 PRELIMINARY; PRT; 253 AA.

AC Q

3

```

GN AF1284.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.W.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RT Nature 390:364-370(1997).
RL EMBL; AE001015; AAB89961.1; -.
DR HSSP; P00274; 2TRX.
DR TIGR; AF1284; -.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRFAMs; TIGR01068; thioredoxin; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 134 AA; 15347 MW; AB8580DC51D86CC2 CRC64;

Query Match      8.9%; Score 120.5; DB 17; Length 134;
Best Local Similarity 31.4%; Pred. NO. 0.00062;
Matches 27; Conservative 19; Mismatches 31; Indels 9; Gaps 3;

QY 104 NDKTIDELELRDKRVTWIWEFFANWSDCOSFAPYADLSLKYNCGLNFGKVDVGRYTD 163
Db 36 NSSNFDEFLKKNENV--VVDWEAECMPCMKIAPVIELAKRY-AGKVYFGKLNTDENPT 92
QY 164 VSTRYKVSTSPSLTKOLPTLILFQGGK 189
  93 IAARYGISA-----IPTLIFPKGK 112

Search completed: July 9, 2003, 11:57:15
Job time : 83 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 12, 2003, 09:52:37 ; Search time 1878 Seconds
(without alignments)
3998.149 Million cell updates/sec

Title: US-09-954-846-2
Perfect score: 1359
Sequence: 1 MAVLAPLIALYVSVPRLSRW.....EQPVASTPTTVDGENKRDK 258

oring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+pn.model -DEV=xlh
-O=/cgn2_1/uspto_spool/US09954846/runat_09072003_102145_12156/app_query.fasta_1.455
-DB=GenEmbl -QWNT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09954846.ecgn_1_1687_runat_09072003_102145_12156 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tabase : GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pt:*
10: gb.to:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*

29: em.vi:*
30: em.htg.hum:*
31: em.htg.inv:*
32: em.htg.other:*
33: em.htg.mus:*
34: em.htg.pln:*
35: em.htg.rod:*
36: em.htg.mam:*
37: em.htg.vrt:*
38: em.sy:*
39: em.htgo.hum:*
40: em.htgo.mus:*
41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1326	97.6	1620	6	AX375964	Sequence
2	1326	97.6	1646	9	BC000666	Homo sapi
3	1326	97.6	1669	9	AF132965	Homo sapi
4	1299	95.6	1631	6	AX136115	Sequence
5	1292	95.1	1644	9	AF059753	Homo sapi
6	1220.5	89.8	1978	10	BC013544	Mus muscu
7	1220.5	89.8	200653	2	AC127695	Mus muscu
8	956	70.3	726	6	AX136467	Sequence
9	920	67.7	189271	9	AL138752	Human DNA
10	867	63.8	112389	2	AC073598	Homo sapi
11	579.5	42.6	110116	9	AP001931	Homo sapi
12	579.5	42.6	158349	2	AC009438	Homo sapi
13	579.5	42.6	161397	2	AC090218	Homo sapi
14	579.5	42.6	170540	2	AC092707	Homo sapi
15	579.5	42.6	199992	2	AC021532	Homo sapi
16	522	38.4	130642	2	AP001082	Homo sapi
17	503.5	37.0	149481	2	AC102678	Mus muscu
18	503.5	37.0	216237	2	AC121786	Mus muscu
19	476.5	35.1	1223	3	AY118488	Drosophil
20	476.5	35.1	31857	2	AC020299	Drosophil
21	476.5	35.1	45105	3	AC004311	Drosophil
22	476.5	35.1	160106	3	AC099014	Drosophil
23	476.5	35.1	173347	3	AC099013	Drosophil
24	476.5	35.1	248096	3	AE003795	Drosophil
25	431.5	31.8	5423	6	AX336311	Sequence
26	431.5	31.8	5423	9	AB002382	Human mrn
27	415	30.5	307	6	AX396397	Sequence
28	336	24.7	117711	9	AP000662	Homo sapi
29	336	24.7	153394	2	AP000727	Homo sapi
30	336	24.7	161397	2	AC090218	Homo sapi
31	331.5	24.4	36882	3	U21324	Caenorhabdi
32	330	24.3	153394	2	AP000727	Homo sapi
33	309	22.7	183656	2	AC111385	Rattus no
34	195	14.3	81994	2	AC096003	Rattus no
35	185.5	13.6	81994	2	AC096003	Rattus no
36	185.5	13.6	183656	2	AC111385	Rattus no
37	138.5	10.2	433	3	AF473536	Schistosom
38	130.5	9.6	1500	6	U01968	Sequence 1
39	130.5	9.6	3552	6	I08049	Sequence 5
40	128.5	9.5	10548	1	AE006828	Sulfolobu
41	127.5	9.4	10733	1	AE013695	Yersinia
42	127.5	9.4	220050	1	AJ414156	Yersinia
43	126.5	9.3	700	8	AY077707	Fusarium
44	126.5	9.3	110000	2	LMFCHK32_22	Continuation (23 o
45	124.5	9.2	1429	3	ACADISPROA	L28174 Acanthamoeb

ALIGNMENTS

```

/organism="Homo sapiens"
/db_xref="LocusID:51075"
/db_xref="taxon:9606"
/clone="MGC:713 IMAGE:3349733"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_l6"

```



```

QY 203 AlaValSerThrProThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
|||||
Db 736 GCTGCTCATGACCTTCTCTGAGGAGATGATCCGAGATTTAACTTAAGTGAAGTGA 795

QY 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal 242
|||||
Db 796 TAGCAGGGGCGCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTGTG 855

QY 243 AlaSerThrProThrValSerAspGlyGluAsnLysLysAspLys 258
|||||
Db 856 GC-TCAACCCCAACAGTGTGATGGGGAACAAGAGGATAAA 902

RESULT 5
AF059753 1644 bp mRNA linear PRI 02-JAN-2001
LOCUS Homo sapiens clone 008a05 My009 protein mRNA, complete cds.
-SESSION AF059753
-VERSION AF059753.1 GI:12001957
KEYWORDS
SOURCE Homo sapiens.
ORGANISM
REFERENCE
AUTHORS Mao, Y.M., Xie, Y., Zhou, Z.X., Ying, K. and Zheng, Z.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Institute of Genetics, School of Life
Science, Fudan University, 220 Handan Rd., Shanghai 200433,
P.R.China

FEATURES
Location/Qualifiers
1..1644
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="008a05"
/tissue_type="brain"
/dev_stage="fetus"
28..1146
/codon_start=1
/product="My009 protein"
/protein_id="AAG43122.1"
/db_xref="GI:12001958"
/translation="MAVLAPLIALIVSVPLRWLAOPYLLSALLSAAPFLVRKLP
LCGLGTORDGNPCDFDREVELMFSLAVMMKRRSITVEOHIGNIFEMFSKVANT
ILFPLDIDRMGLLYITLCIVELMTCKPPLYMGPEYIKYFNDKTIDELEDRKRTWIV
EPFANNSDCOSAPFIADLSLYNCTGLNFGKVDVGRYTDVSTRYKVSPLTKQLP
TLILFGGKEAMRRPOISKRGRAVSWTFSEENVIREFNELNELYORAKKLKAGDNIEP
EQPVASTPQCCOMKTRRNKILTLAVLPLLSIPGSPHNHKEPAAAFYLCFPFGCDWV
GOHAASDFKEASRELSGTLOEGLPCCQOLFWSKKEIS"
BASE COUNT 412 a 401 c 385 g 446 t
ORIGIN

Alignment Scores:
Pred. No.: 1.73e-137 Length: 1644
Score: 1292.00 Matches: 255
Percent Similarity: 86.49% Conservative: 1
Best Local Similarity: 86.15% Mismatches: 2
Query Match: 95.07% Indels: 39
DB: 9 Gaps: 1

US-09-954-846-2 (1-258) x AF059753 (1-1644)

QY 1 MetAlaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp 20
|||||
Db 28 ATGGCGGCTTGGCACTCTAAATGCTCTGCTATTCGGTGGCGGCACTTTCACGATGG 87

QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuSerAlaAlaPheLeuValArg 40
|||||
Db 88 CTGCCCCAACCTTACTTACCTTCTGTCGGCCCTGCTCTGCTGCTTCTTCTGCTGAGG 147

QY 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
|||||
Db 148 AAACGCGCGGCTCTGCCACGCTCTGCCACCCACGCGGCAAGACGGTACCCGCTGTGAC 207

```

```

QY 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
|||||
Db 208 TTTGACTCGAGAGAAGTGGAGATCTCTGATGTTTTCAGTGCCATTGTGATGATGAAGAAC 267

QY 81 ArgArgSer----- 83
|||||
Db 268 CGCAGATCCATCACTGTGGAGCAACATATAGSCAACATTTTCATGTTTAGTAAAGTGCC 327

QY 83 ----- 83
|||||
Db 328 AACACAATTCTTTTCTCCGCTTGGATATTCGCATGGCCCTACTTTACATCACACTCTGC 387

QY 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
|||||
Db 388 ATAGTGTCTCGATGACGTCAAAACCCCTATATATGGGCCCTGAGTATATCAAGTAC 447

QY 103 PheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrpIleVal 122
|||||
Db 448 TTCAATGATAAAACCATTTGATGAGCACTAGAACGGGACAGAGGCTCACTTGGATTGTG 507

QY 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
|||||
Db 508 GAGTCTTTGGCAATTGCTTAATGACTGCCAATCATTTGCCCTCTATCTATGCTGACCTC 567

QY 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
|||||
Db 568 TCCCTTAATAACACTGTACAGGGCTAAATTTTGGGAAGGTGGATTTGGAGCGCTATACT 627

QY 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
|||||
Db 628 GATGTTAGTAGCGGTACAAAGTGACACATCACCCCTCACCAACAACCTCCTACCCTG 687

QY 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
|||||
Db 688 ATCCTGTTCCAAGTGGCAAGGAGCAATCGCGGCGCACAGATTAGCAAGAAGGCGG 747

QY 203 AlaValSerThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
|||||
Db 748 GCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCCGAGAAATTTAACTTAATGAGCTA 807

QY 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProVal 242
|||||
Db 808 TACCAGCGGCGCAGAACTATCAAAAGCTGGAGACAATATCCCTGAGGAGCAGCTGTG 867

QY 243 AlaSerThrProThrValSerAspGlyGluAsnLysLysAspLys 258
|||||
Db 868 GCTTCAAC-CCCAACACAGTCTCAGATGGGGAACAAGAGGATAAA 914

RESULT 6
BC013544 1978 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, RIKEN cDNA 2310042M24 gene, clone MGC:19231
DEFINITION IMAGE:4242418, mRNA, complete cds.
ACCESSION BC013544
VERSION BC013544.1 GI:15488826
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1978)
Direct Submission
Submitted (04-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNI)

```

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 25 Row: b Column: 24
 This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction, similarity but not identity to protein.

FEATURES

source

1. 1978
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:19231 IMAGE:4242418"
 /tissue_type="Kidney, normal, 5 month old male, mouse."
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

CDS

4. 891
 /codon_start=1
 /product="RIKEN cDNA 2310042M24 gene"
 /protein_id="AAH13544.1"
 /db_xref="GI:15488827"
 /db_xref="LocusID:66958"
 /translation="MAVLAPLIALIYSVPRLSRWLRALPYCLLSALLSTAFLLVRLKLP
 ICGLPLORENGPCDFDWEVEILMFSAIVMMKNRSITVEQHVGNLFMFSSKANA
 ILFPLDRLMGLLYLTICIVELMTCKPPLYMGPEYIKYENDKTIDEELERDKRWIV
 EFFANWSDKSFARIDYADLSKYNCSGLNFKGVDRITSTRYKYSSTSLTQLP
 TLILFQKGVIRPOIDKGRVSWTFSEENVIREFNELNYQRAKKHSGKGDWSEE
 KVPDPAPTVPDGENKKDK"

BASE COUNT 509 a 454 c 472 g 543 t

ORIGIN

Alignment Scores:
 Pred. No.: 3,17e-129 Length: 1978
 Score: 1220.50 Matches: 238
 Percent Similarity: 83.45% Conservative: 10
 Best Local Similarity: 80.41% Mismatches: 39
 Query Match: 89.81% Indels: 2
 DB: 10 Gaps: 2

US-09-954-846-2 (1-250) x BC013544 (1-1978)

Qy 1 MetAlaValLeuAlaProLeuLeuAlaLeuValTyrSerValProArgLeuSerArgTtp 20
 Db 4 ATGGCTGTCTTGGCCCTCTGATGCTTGGTGTACTCGTGGCCGGCTTCTCGATGG 63
 Db 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
 Qy 64 CTGGCCCGACCTATTGCTCTGCTGCTGCTTCCATTGCTTCTCTCTCGTGAGG 123
 Db 41 LysLeuProProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 Db 124 AAATGCCACCGATTGCAATGGTCTCCACGCAACGCAAGATGGCAACCCGTTGAC 183
 Qy 61 PheAspTtpArgGluValGluLeuLeuMetPheLeuSerAlaLeuValMetMetLysAsn 80
 Db 184 TTGACTGGAGAAAGTGGATCTCTGATGTTCTCAGTGCCATTGTGATGATGAAGAAC 243
 Qy 81 ArgArgSer----- 83
 Db 244 CGCAGATCCATCAGTGTGGAGCAACATGATGAGCAACATCTTATGTTTAGTAAGTGGCC 303
 Qy 83 ----- 83
 Db 304 AACGCCATCTTTCTCCGACTGGATATTGCAATGGGTCTGCTATACCTCAGCTCTGC 363
 Qy 84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102

Db 364 ATAGTGTCTTCTGATGACCTGCAAGCCCGCTGTACATGGTCTCAGTATATCAAGTAC 423
 Qy 103 PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrIleVal 122
 Db 424 TTCAATGATATAAACCATTTGATGAGGAGCTGGAGCGAGACAAGAGGCTCATTTGGATTGTG 483
 Qy 123 GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
 Db 484 GAGTCTTTTGGCAACTGGTCTATGATGCTCCATCTTTGCTCCCATCTATGCGGACTTG 543
 Qy 143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
 Db 544 TCCCTCAACTACACTGTTTTCAGGGCTAAATTTTGGCAAGGTAGATGTTGGACCTACACT 603
 Qy 163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
 Db 604 GACGTTAGCACACGGTACAAAGTGAGCACATCACCCCTCACCAGACAGCTCCCTACCCCTG 663
 Qy 183 IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg 202
 Db 664 ATTCTGTTCCAAAGCGCGAAGAGGTCTATCGTCGGCCGAGATTGACAAAGAGACGA 723
 Qy 203 AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu 222
 Db 724 GCTGCTCTTGGACCTTTTCTGAGGAGATGTTGATTCGAGATTCAACTTGAATGAGCTA 783
 Qy 223 TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGlnProVal 242
 Db 784 TACCAACGAGCAAGCAAGACACTCAAAGGCTGGAGAC---ATGTGCAAGAAGAAGCCTGTG 840
 Qy 243 AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
 Db 841 GACCCTGCTCCCACTACTGTGCGAGATGGGAAACAAAGAAGGACAAA 888

RESULT 7
 AC127695 200653 bp DNA linear HTG 11-AUG-2002
 LOCUS Mus musculus chromosome UNK clone RP23-310B17, WORKING DRAFT
 DEFINITION AC127695
 AC127695
 AC127695
 VERSION AC127695.2 GI:22203922
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1. (bases 1 to 200653)
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 2 (bases 1 to 200653)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (18-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 200653)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (11-AUG-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 On Aug 11, 2002 this sequence version replaced gi:21903674.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: M_BA0310B17
 ----- Summary Statistics -----
 Sequencing, vector: M13; 0%

Qy	103	PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrTrpIleVal	122
Db	126328	TTCAATGATAAAACCATTTGATGAGGAGCTGGAGCGAGACAAGAGGCTCACTTGGATTGTG	126387
Qy	123	GluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrrAlaAspLeu	142
Db	126388	GAGTTCCTTGGCAACTGGTCTAAATGATGGCCAGTCCCTTTGGTCCCATCTATCGGCACTTG	126447
Qy	143	SerLeuLysTyrrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr	162
Db	126448	TCCCTCAAGTACAACTGTTCAAGGCTAAATTTTGGGAAGGTAGATGTTGGACGCTACACT	126507
Qy	163	AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu	182
Db	126508	GACGTTAGCACACGGTAACAAAGTAAGACACATCACCCCTCACCCAGACAGCTCCCTACCCCTG	126567
Qy	183	IleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArg	202
Db	126568	ATTCGTGTTCCAAGCGCGCAAGAGGTCAATTCGTGGCGCGCAGATTGCACAAGAAAGGACGA	126627
Qy	203	AlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsnGluLeu	222
Db	126628	GCTGTCCTTGGACCTTTCTCGAAGAGAATGTGATTCAGAAATTCAACTTGAATGAGCTA	126687
Qy	223	TyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGlnProVal	242
Db	126688	TACCAACGAGCGCAAGAAGCACTCAAAAGGGTGGAGAC---ATGTCAGAAGAAGAGCTGTG	126744
Qy	243	AlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys	258
Db	126745	GACCTGCTCCCACTACTGTGCCAGATGGGAAAACAAGAAGGACAAA	126792
RESULT 8			
AX136467			
LOCUS	AX136467	726 bp	DNA linear PAT 30-MAY-2001
DEFINITION	Sequence 389 from Patent EP1067182.		
ACCESSION	AX136467		
VERSION	AX136467.1 GI:14272871		
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and Hayashi, K.		
TITLE	Secretory protein or membrane protein		
JOURNAL	Patent: EP 1067182-A 389 10-JAN-2001;		
FEATURES	Helix Research Institute (JP)		
source	Location/Qualifiers		
	1..726		
BASE COUNT	176 a	191 c	190 t
ORIGIN	8 others		
Alignment Scores:			
Pred. No.:	1.4e-99	Length:	726
Score:	956.00	Matches:	188
Percent Similarity:	80.59%	Conservative:	3
Best Local Similarity:	79.32%	Mismatches:	7
Query Match:	70.35%	Indels:	39
DB:	6	Gaps:	1
US-09-954-846-2 (1-258) x AX136467 (1-726)			
Qy	1	MetaValLeuAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp	20
Db	16	ATGCGGCTGTGGCACTCTAAATTCGTCTCGTGTATTGGTCCGCGCACTTTCACGATGG	75
Qy	21	LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuSerAlaAlaPheLeuValArg	40

```
Db      76 CTCGCCCAACCTTACTACTCTGTGCGCCCTCTCTCTGTGCGCTTCTACTCTGTGAGG 135
Oy      41 LysLeuProLeuCyHisGlyLeuProThrGlnAtgGluAspGlyAsnProCysasp 60
Db      136 AAATGCCCGCCCTCTGACCGGTCTGCCACCAACGAGAGCGGTAAACCCGGTGTGAC 195
Oy      61 PheAspTrpArgGluValIleLeuMetPheLeuSerAlaIleValIleMetMetLysAsn 80
Db      196 TTTGACTGGAGAGAGTGAGATCCTGATGTTTCTCAGTGCCTGATGATGATGAAGAC 255
Oy      81 ArgArgSer----- 83
Db      256 CGCAGATCCACTGCTGTGGAGCAACATATAGGCAACATTTTCATGTTTAGTAAAGTGCC 315
Oy      83 ----- 83
Db      316 AACACAATCTTTCTTCGCTTGATATATCGCATGGCGCTACTTTACATCATCTCTGC 375
Oy      84 ---MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyr 102
Db      376 ATAGTGTTCTGTGATGACGTGCAACACCCCTATATATATGGCCCTGAGTATATCAAGTAC 435
Oy      103 PheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrTrpIleVal 122
Db      436 TTCAATGATNAACCATTCATGATGAGGAAGTACAGCGGACAGGAGGTCACCTGGATTGTG 495
Oy      123 GluPhePheAlaAsnTyrSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeu 142
Db      496 GAGTCTTTTGCAATGTGTCTAATGACTGCGCAATCATTTGCCCCCTATCTATCTGACCTC 555
Oy      143 SerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArgTyrThr 162
Db      556 TCCCTTAATACAACTGTACAGGGCTAATTTGGGAANGTGGATGTINGACGCTATACT 615
Oy      163 AspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeu 182
Db      616 GATGTTAGTACGGGTACAAATGAGCACATCACCCCTCACCAAGCAACTCCCTACCCCTG 675
Oy      183 IleLeuPheGlnGlyLysGluAlaMet-ArgArgProGlnIleAsp 198
Db      676 ATNCTGTTCAGGTGGCAAGCAATGCGNGCGCCCNNAATTGAC 724

RESULT 9
LOCUS   AL138752/c
DEFINITION Human DNA sequence from clone RP11-3J10 on chromosome 9p12-13.3
          Contains part of a novel gene, the 3' end of the SHB (SHB adaptor
          protein (a Src homology 2 ) protein ) gene, the gene for CG7943
          protein, a pseudogene similar to CGI-31 protein, a pseudogene
          similar to PAICS (phosphoribosylaminimidazole carboxylase,
          phosphoribosylaminimidazole succinocarboxamide synthetase), part
          of a gene for a ribosomal protein, the 3' end of gene for
          apoptosis-related protein PNAS-3, ESTs, GSSs and CpG islands,
          complete sequence.
ACCESSION AL138752
VERSION   AL138752.5 GI:8452480
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189271)
AUTHORS   Laird G.
TITLE     Direct Submission
JOURNAL   Submitted (08-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
          requests: clonerquest@sanger.ac.uk
COMMENT   On Jun 10, 2000 this sequence version replaced gi:8250065.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
```

only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>

This sequence is the entire insert of clone RP11-3J10 This sequence has been finished according to sequence map criteria as follows.

An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

RP11-3J10 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

Location/Qualifiers

source

1..189271

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="p12-13.3"

/clone="RP11-3J10"

/clone_lib="RPCI-11.1"

126..493

/note="12 repeat: matches 2325..2705 of consensus"

1456..1537

/note="MER66B repeat: matches 271..349 of consensus"

MER66A repeat: matches 263..341 of consensus"

2434..2532

/note="MIR repeat: matches 23..122 of consensus"

3695..3917

/note="MIR repeat: matches 9..256 of consensus"

4195..4263

/note="MIR repeat: matches 83..152 of consensus"

5058..5122

/note="MIR repeat: matches 39..103 of consensus"

5751..5854

/note="MIR repeat: matches 65..169 of consensus"

7967..8173

/note="12 repeat: matches 2457..2687 of consensus"

8421..8481

/note="07 repeat: matches 1..62 of consensus"

U7 repeat: matches 1..62 of consensus"

8638..8687

/note="12 repeat: matches 2645..2696 of consensus"

8948..9233

/note="AluX repeat: matches 36..308 of consensus"

9410..9571

/note="MIR repeat: matches 5..190 of consensus"

10596..10867

/note="AluJo repeat: matches 8..276 of consensus"

11536..11593

/note="MER5A repeat: matches 128..186 of consensus"

11937..11971

/note="MIR repeat: matches 112..146 of consensus"

12275..12551

/note="AluJo repeat: matches 1..274 of consensus"

12555..12885

/note="L1MA9 repeat: matches 5880..6216 of consensus"

L1MA9 repeat: matches 5880..6216 of consensus"

12886..13189

/note="AluSg repeat: matches 1..313 of consensus"

13190..14503

/note="L1MA9 repeat: matches 4579..5880 of consensus"

L1MA9 repeat: matches 4579..5880 of consensus"

14504..15045

/gene="BA3J10.1"

<14504..15045

mRNA


```

Db 59008 AAGTAATCGGAGCAGACAGACTGACAAAGAGCGGCTGCTCATGACCTTCTCT 58949
QY 210 AGUGluasnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArGAlaLysLysLeu 229
Db 58948 GAGGAGAATGTGATCCGAGAATTAACTTAATGAGCTATACACGGCGGCAAGAGCTA 58889
QY 230 SerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThrProThrVal 249
Db 58888 TCAAGGCTGGAGACATATCCCTGAGGAGCAGCCTGTGGCTTCAACCCCAACACAGTG 58829
QY 250 SerAspGlyGluAsnLysLysAspLys 258
Db 58828 TCAGATGGGGAACACAGAGGATAAA 58802

RESULT 10
AC073598 Homo sapiens chromosome 10 clone CTC-268N23, WORKING DRAFT
LOCUS AC073598
DEFINITION AC073598.3 GI:20279338
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
~URCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 112389)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.B., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs1,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S.,
Karissom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mathiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,K.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 112389)
Worley,K.C.
Direct Submission

```

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BASE COUNT 29689 a 25997 c 25893 g 29967 t 843 others
ORIGIN
Alignment Scores:
Pred. No.: 1.17e-86 Length: 112389
Score: 867.00 Matches: 181
Percent Similarity: 74.70% Conservative: 5
Best Local Similarity: 72.69% Mismatches: 19
Query Match: 63.80% Indels: 45
DB: 2 Gaps: 3
US-09-954-846-2 (1-258) x AC073598 (1-112389)

```

```

Submitted (26-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 112389)
Worley,K.C.
Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 24, 2002 this sequence version replaced gi:16118028.
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: HMRT
Center clone name: CTC-268N23
-----
Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 10470 bases at least Q40
Consensus quality: 107539 bases at least Q30
Consensus quality: 109442 bases at least Q20
Estimated insert size: 111011; sum-of-ctigs estimation
Quality coverage: 3x in Q20 bases; sum-of-ctigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank.draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2191: contig of 2191 bp in length
* 2192 2291: gap of unknown length
* 2292 5898: contig of 3607 bp in length
* 5899 5998: gap of unknown length
* 5999 11406: contig of 5408 bp in length
* 11407 11506: gap of unknown length
* 11507 21739: contig of 10233 bp in length
* 21740 31956: contig of 10117 bp in length
* 31957 32056: gap of unknown length
* 32057 46774: contig of 14718 bp in length
* 46775 46875: gap of unknown length
* 46876 68219: contig of 21344 bp in length
* 68219 68318: gap of unknown length
* 68319 85954: contig of 17636 bp in length
* 85955 86054: gap of unknown length
* 86055 112389: contig of 26335 bp in length.
*
* Location/Qualifiers
* 1. 112389
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="10"
* /clone="CTC-268N23"
*
BASE COUNT 29689 a 25997 c 25893 g 29967 t 843 others
ORIGIN
Alignment Scores:
Pred. No.: 1.17e-86 Length: 112389
Score: 867.00 Matches: 181
Percent Similarity: 74.70% Conservative: 5
Best Local Similarity: 72.69% Mismatches: 19
Query Match: 63.80% Indels: 45
DB: 2 Gaps: 3
US-09-954-846-2 (1-258) x AC073598 (1-112389)

```



```
QY 53 ArgGluAspClyAsnPro-----CysAspPheAspTrpArgGluValGluIleLeu 69
|||||
Db 19522 AGAAGACTCGGAACCCCAAGAGTGTGCCGAACCGGAGAGGAGGAGATCCTG 19581
QY 70 MetPheLeuSerAlaIleValMetMetLysAsnArgSer----- 83
|||||
Db 19582 ATGTTTCTCAGTCGCATTTGTGATGATGAAGAACCAGCATCCACTCTGTGGAGCAACAT 19641
QY 83 ----- 83
Db 19642 ATAGGCAACATTTTCATGTTTAGTAAAGTGGCAACGCAATCTTTTCTCCGCTGGAT 19701
QY 84 -----MetPheLeuMetThrCysLys--- 90
|||||
19702 ATTCCGATGGGCTACTTTACATCACACTCTGCATAGTGTCTCTGATGAGCGTCAACACC 19761
QY 91 ---ProProLeuThrMetGlyProGluTyrIleLysTyrPheAsnAspLysThrIleAsp 109
|||||
Db 19762 CCCCTCCCA-TATATGGCCCTGAGTATATCAAGTACTTCAATGATAAACCATTGAT 19820
QY 110 GluGluLeuGluArgAspLysArgValThrTrpIleValGluPheAlaAsnTrpSer 129
|||||
Db 19821 GAGGAACCTAGAACGGGACACACAGGTCACCTGGATTTGGAGTCTTTGCCAATTGGTCT 19880
QY 130 AsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLysTyrAsnCysThr 149
|||||
Db 19881 AATGACTGCCAATCATTTGCCCTATCTATGCTGACCTCTCCCTTAAGTACAACACTGACA 19940
QY 150 GlyLeuAsnPheGlyLysValAspValGlyArgTyrThrAspValSerThrArgTyrLys 169
|||||
Db 19941 GGGCAAAATTTGGGACGGTGGATGCTGGACGCTACTAGTGTAGTACCGGTACAAA 20000
QY 170 ValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGlnGlyLys 189
|||||
Db 20001 GTGAGCACATCACCCCTCACCAATCACTCCCTACCTGATCTGTTCGAAGTGGCAAG 20060
QY 190 GluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSerTrpThr-PheSe 209
|||||
Db 20061 AAGGTAATCGCGGACGACAGACTGANCAGAAGGACGGGCTGTCTCATGGACCTTTCTC 20120
QY 209 rGluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAlaLysLysLe 229
|||||
Db 20121 TGAGGAGATGTGATCCGAGATTTAACTTAATGAGCTATACACGCGGGCAGAGCT 20180
QY 229 uSerLysAlaGlyAspAsnIleProGluGlnProValAlaSerThrProThrTrpVa 249
|||||
Db 20181 TTCAAAGGCTGGAGACAATATCCCTGAGGAGCAGGCTGGGCTTTCAGCCCCCACCACAGT 20240
QY 249 lSerAspGlyGluAsnLysLysAsp 257
|||||
Db 20241 GTCAGATGGGAAAGCAAGAGGAT 20265

RESULT 11
AP001931
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 11q clone:RP11-691N7, complete
ACCESSION AP001931
VERSION AP001931.5 GI:21218128
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
1 Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
  Toki,I., Watanabe,H. and Sakaki,Y.
  Published Only in Database (2000)
  2 (bases 1 to 110116)
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Toki,I., Watanabe,H. and Sakaki,Y.

TITLE
JOURNAL
COMMENT On May 27, 2002 this sequence version replaced gi:20334325.
FEATURES
Location/Qualifiers
source
1..110116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-691N7"
BASE COUNT 29082 a 23179 c 24671 g 33184 t
ORIGIN
Alignment Scores:
Pred. No.: 7,5e-54 Length: 110116
Score: 579.50 Matches: 172
Percent Similarity: 27.37% Conservative: 1
Best Local Similarity: 27.22% Mismatches: 2
Query Match: 42.64% Indels: 459
DB: 9 Gaps: 4

US-09-954-846-2 (1-258) x AP001931 (1-110116)
QY 84 MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTrpPhe 103
|||||
Db 18410 GTCTTCTGATGAGTGCACAAACCCCTATATATGGCCCTGAGTATATCAAGTACTTC 18469
QY 104 AsnAspLysThrIleAspGlu----- 110
|||||
Db 18470 AATGATAAAACCATTTGATGT-GAGTGCTCTTTCCCTTTCTGTTCTTGGTCCCTTGTG 18528
QY 110 ----- 110
Db 18529 GGTGATTTTGTAGTTGTGCTCTCCATCTACTAGGAGAACACAGCTCTTCAAAATGAT 18588
QY 110 ----- 110
Db 18589 GTCACGGGCACTGTGTTTTCATTATGAGAGCCGGGAGAGGAGACACAGGATCAGGAT 18648
QY 110 ----- 110
Db 18649 TAGATGCTAAAGTCTGAACCTTCCAGGAGGCTATGTGGATCCAGCTGACTTTTCTTCCC 18708
QY 111 -----GluLeuGluArgAspLysArgValThrTrpIleValGluPhePhe 125
|||||
Db 18709 TGTATTGGCAGGAGGAACACTAGAACGGGACAGAGGCTCATTGGATTGTGGAGTCTTT 18768
QY 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
|||||
Db 18769 GCCAATGGTCTAATGACTGCCAATCATTTGCCCTTATCTATCTGCTGACCTCTCCCTTAA- 18827
QY 145 ----- 145
Db 18828 GTGAGTAGTGCAAAAGGAGGAGGATGGTGGAAATGGAGATGCTGTGCCTTCCCTCTCACTGT 18887
QY 145 ----- 145
Db 18888 TTTTGGCTTTTCTTTTCTTTTCTTTTGGCCTTGATTTTTCACACATGGTAAACCAAGCATCTCC 18947
QY 145 ----- 145
Db 18948 CTCCTCCCTCTTAAATATCTATACTTCCACTTTCTTGTGATCCATATATTTTTCAGCAT 19007
QY 146 -----TyrAsnCysThrGlyLeuAsnPheGlyLysValAsp 157
|||||
Db 19008 ATTAATAATATATTTCTTTTTCAGATACAACTGTACAGGGCTAAATTTTGGGAAGGTGAT 19067
QY 158 ValGlyArgTyrThrAspValSerThr----- 166
|||||
```

Db 19068 GTTGGAGCTTACTATGATGTTAGTACCGGTATGTAAAGACCTGGCGAGAGGGTCTGAGC 19127
QY 167 -----Ar 167
Db 19128 AGGGAATCACCTTTGAGTGATACATACATACAGGACATTTTAGAGAACTTTCTGGGCCCTGCAG 19187
QY 167 gTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuLeuLeuPheGlnG1 187
Db 19188 GTACAAGTGAGCACATFCACCCCTCACAAGCACTCCCTACCTGATCTCTGTTCACAGG 19247
QY 187 yGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSerTrpTh 207
Db 19248 TGGCAAGGAGGCAATCGCGGGCCACAGATTTGACAAAGAGGCGCTGTCTCATGGAC 19307
QY 207 rPheSerGlu-----210
Db 19308 CTTCTCTGAGGTACCTGAAAGGAAGGCGCAGTGCGATGAAGGGTGCGAAGCAGTAGGTGGG 19367
QY 210 -----210
Db 19368 CTTTCAAGCCCTACCCGGGTTTGATTCACAGCTCTGCCACTTGGCCCAATTAGCTTTGAGGG 19427
QY 210 -----210
Db 19428 TGTGGACTACTTACTAGACCTCATGTTTAAATTCATTAGCAATATATATTATGTGCAG 19487
QY 210 -----210
Db 19488 ACCTGCTAGTACTGGGTAAACAATCATTTAGTTAAAGATTGGTCTCAGGCCAGCGGTGG 19547
QY 210 -----210
Db 19548 CTCACCCCTGTAATCCAGCACCTTTGGGAGCGCGAGCTGGCAGATCACCTGAGGTGAGG 19607
QY 210 -----210
Db 19608 AGTTTGAGACCGCTGACCAACATGAGAGAAACCCCGTGTCTACTAAATAACAATA 19667
QY 210 -----210
Db 19668 GCCGAGTGTGGTGGCGCATGCTATATATGCCAGTATTTGGGAGGCTGAGGAGGAGAA 19727
QY 210 -----210
Db 19728 CGCTTGAACCCGGAGGTGGAGTTGCGAGTGAGTGCACCATTTGGGTTCCAGCC 19787
QY 210 -----210
Db 19788 TGGGCAACAAGAGACTCCATCAGGAAAAAAGAAAGATTCTGTCTCGAAGGCCACTGC 19847
QY 210 -----210
Db 19848 ACCTATTGTCTGTAGGTGTAGGTTCCTGCTCAGTTAAATTGGAGGTAATAACCCCTGGCT 19907
QY 210 -----210
Db 19908 TACAGGGTGTGTTGCGAGATTACATTAATAATTAGGTATATCTTGGCAGCGGTGATT 19967
QY 210 -----210
Db 19968 GGCACATAATGGCTATTTAATACTATTAGTTGACTTTTCTGGGTACCTAAAGAGAGG 20027
QY 210 -----210
Db 20028 TTAGGAAGATTTTGTCTCTTGTCTTACTCTCTCTTCCAGACTTTGTGTAAATACC 20087
QY 210 -----210
Db 20088 TCTTACTTCCAGGCTTTTACTCTCCCTTCCAAACCCAGATCTCTGAGTGTGCATCTCT 20147
QY 211 -----GluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArqAlaLysL 228
Db 20148 TTCTGCGAGAGAAATGTGATCCGAGAAATTTAACTTAATAGCTATACACGGGGCCAGA 20207

QY 228 ysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThrProThrT 248
Db 20208 AACATATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCA 20267
QY 248 hrValSerAspGlyGluAsnLysLysAspLys 258
Db 20268 CAGTGTCAAGTGGGAAACACAGAGGATAAA 20299

AC009438 158349 bp DNA linear HTG 04-SEP-2000
Homo sapiens chromosome 11 clone RP11-77M17 map 11, WORKING DRAFT
SEQUENCE, 25 unordered pieces.

AC009438
AC009438.3 GI:9966254
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens.
Homo sapiens.

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 158349)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Unpublished

TITLE
JOURNAL
AUTHORS

2 (bases 1 to 158349)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
Funke, R., Gage, B., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, B., Howland, J. C., Jones, C., Kann, L.,
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission

Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7655447.
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1359

Center clone name: 77_M17

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-primer-amersham; 5% of reads

Chemistry: Dye-terminator Big Dye; 95% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 138242 bases at least Q40

Consensus quality: 145728 bases at least Q30

Consensus quality: 149650 bases at least Q20

Insert size: 172000; agarose-fp

Insert size: 155949; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

* 1
* 3650: contig of 3650 bp in length
* 3651 3750: gap of 100 bp
* 3751 4867: contig of 1117 bp in length
* 4868 4967: gap of 100 bp
* 4968 6757: contig of 1790 bp in length
* 6758 6857: gap of 100 bp
* 6858 7991: contig of 1134 bp in length
* 7992 8091: gap of 100 bp
* 8092 9216: contig of 1125 bp in length
* 9217 9316: gap of 100 bp
* 9317 10384: contig of 1068 bp in length
* 10385 10484: gap of 100 bp
* 10485 11835: contig of 1351 bp in length
* 11836 11935: gap of 100 bp
* 11936 13429: contig of 1494 bp in length
* 13430 13529: gap of 100 bp
* 13530 15092: contig of 1563 bp in length
* 15093 15192: gap of 100 bp
* 15193 17380: contig of 2188 bp in length
* 17381 17480: gap of 100 bp
* 17481 20586: contig of 3106 bp in length
* 20587 20686: gap of 100 bp
* 20687 24958: contig of 4272 bp in length
* 24959 25058: gap of 100 bp
* 25059 30271: contig of 5213 bp in length
* 30272 30371: gap of 100 bp
* 30372 35676: contig of 5305 bp in length
* 35677 35776: gap of 100 bp
* 35777 40640: contig of 4864 bp in length
* 40641 40740: gap of 100 bp
* 40741 46368: contig of 5628 bp in length
* 46369 46468: gap of 100 bp
* 46469 53373: contig of 6905 bp in length
* 53374 53473: gap of 100 bp
* 53474 60571: contig of 7098 bp in length
* 60572 60671: gap of 100 bp
* 60672 67052: contig of 6381 bp in length
* 67053 67152: gap of 100 bp
* 67153 75831: contig of 8679 bp in length
* 75832 75931: gap of 100 bp
* 75932 84118: contig of 8187 bp in length
* 84119 84218: gap of 100 bp
* 84219 95075: contig of 10857 bp in length
* 95076 95175: gap of 100 bp
* 95176 110669: contig of 15494 bp in length
* 110670 110769: gap of 100 bp
* 110770 139814: contig of 29045 bp in length
* 139815 139914: gap of 100 bp
* 139915 158349: contig of 18435 bp in length.
  
```

FEATURES

source

```

1..158349
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /map="11"
  /clone="RP11-77M17"
  /clone_lib="RPC1-11 Human Male BAC"
misc_feature
  1..3650
  /note="assembly_fragment"
  clone_end:Sp6
  vector_side:left
misc_feature
  3751..4867
  /note="assembly_fragment"
misc_feature
  4968..6757
  /note="assembly_fragment"
misc_feature
  6858..7991
  /note="assembly_fragment"
misc_feature
  8092..9216
  /note="assembly_fragment"
misc_feature
  9317..10384
  /note="assembly_fragment"
  
```

```

misc_feature
  /note="assembly_fragment"
  10485..11835
misc_feature
  /note="assembly_fragment"
  11936..13429
misc_feature
  /note="assembly_fragment"
  13530..15092
misc_feature
  /note="assembly_fragment"
  15193..17380
misc_feature
  /note="assembly_fragment"
  17481..20586
misc_feature
  /note="assembly_fragment"
  20687..24958
misc_feature
  /note="assembly_fragment"
  25059..30271
misc_feature
  /note="assembly_fragment"
  30372..35676
misc_feature
  /note="assembly_fragment"
  35777..40640
misc_feature
  /note="assembly_fragment"
  40741..46368
misc_feature
  /note="assembly_fragment"
  46469..53373
misc_feature
  /note="assembly_fragment"
  53474..60571
misc_feature
  /note="assembly_fragment"
  60672..67052
misc_feature
  /note="assembly_fragment"
  67153..75831
misc_feature
  /note="assembly_fragment"
  75932..84118
misc_feature
  /note="assembly_fragment"
  84219..95075
misc_feature
  /note="assembly_fragment"
  95176..110669
misc_feature
  /note="assembly_fragment"
  110770..139814
misc_feature
  /note="assembly_fragment"
  139915..158349
misc_feature
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right
  
```

BASE COUNT 40315 a 36712 c 36032 g 42881 t 2409 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1.19e-53 Length: 158349
 Score: 579.50 Matches: 172
 Percent Similarity: 27.37% Conservative: 1
 Best Local Similarity: 27.22% Mismatches: 2
 Query Match: 42.64% Indels: 459
 DB: 2 Gaps: 4

US-09-954-846-2 (1-258) x AC009438 (1-158349)

```

QY      84 MetPheLeuMetThrCysLysProProLeuTyMetClyProGluTyrIleLysTyPhe 103
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      143071 GTGTTCTTGATGACGTGCAAAACCCCTATATATGGGCCCTGAGTATATCAAGTACTTC 143130
       |||
QY      104 AsnAspLysThrIleAspGlu-----
       |||
Db      143131 AATGATAAAACCAATGATGT-GAGTGTCTTTCCCTTTCTGTGTTCTTGGTCCCTGTG 143189
       |||
QY      110 -----
       |||
Db      143190 GGTGATTTTGACTGTGCTCTCCATTCAGGAGGAACAACAGTCTTCAAAATGGAT 143249
       |||
QY      110 -----
       |||
Db      143250 GTCACGGGCACTGTGTTTCATTATGAGACCGGGGAGAGGAACACAGGATGAGGAAT 143309
       |||
QY      110 -----
       |||
Db      143310 TAGATCTAAAGTCTGAACCTTCCCAGGAGGCTATGTGGATCCAGCTGACTTTCTTCTCC 143369
       |||
  
```

```
QY 111 -----GluLeuGluArgAspLysArgValThrPheIleValGluPhePhe 125
Db 143370 TGTATTTGGCAGGAGGAACTAGACGGACAAAGAGGTCACTTGGATTGTGGAGTCTTT 143429
QY 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
Db 143430 GCCAATTTGGTCTAATGACTGCGCAATCATTTGCCCTATCTATGCTGACCTCTCCCTTAA- 143488
QY 145 ----- 145
Db 143489 GTGAGTAGTCAAAAGGGGAGGTGGAAATGGAGATGCTGTGCCCTCCCTCTCACGTGT 143548
QY 145 ----- 145
Db 143549 TTTTGGCTTTTCTTTTCTTTTGGCCTTGATTTTTCACACATGGTAACCAAGGCATCTCC 143608
QY 145 ----- 145
Db 143609 CTCTCCCTCTTAAATATCTATCTTCCACTTTCTTGTGATCCATTAATTTTTTTCAGCAT 143668
QY 146 -----TyrAsnCysThrGlyLeuAsnPheGlyLysValAsp 157
Db 143669 ATTAATAATATATTTCTTTTCAGATACAACTGTACAGGCTAAATTTTGGGAAGTGGAT 143728
QY 158 ValGlyArgTyrThrAspValSerThr----- 166
Db 143729 GTTGGACGCTATACTGATGTGTAGTACGCGTATGTAAGACCTGGCGAGAGGCTCTGAGC 143788
QY 167 -----Ar 167
Db 143789 AGGAAATCACTTTGAGTGATACATACAGGGACATTTAGAGAACTTTCTGGGCCCTGCAG 143848
QY 167 qTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGlnG1 187
Db 143849 GTCAAAGTGAGCATACACCCCTCACCAACCACTCCCTACCTGATCTCTGTTCCAAAG 143908
QY 187 yGlyLysGluAlaMetArgArgProGlnIleAspLysGlyArgAlaValSerTrpTh 207
Db 143909 TGGCAAGGAGGCAATGGCGGCCACAGATTGACAAGAAGAGCGGGTCTCTCATGGAC 143968
QY 207 rPheSerGlu----- 210
Db 143969 CTCTCTGAGGTACTGAAAGGAGGGCAGGTGCATGAGGTGCAGAACAGTAGGTGG 144028
QY 210 ----- 210
Db 144029 CTTTCAAGCCCTACCCGGTTTGATTACAGCTCTGCCACTTGCCCATTTAGCTTTGAGGG 144088
QY 210 ----- 210
Db 144089 TGTGGACTAGTTACTAGACCTCATGTTTAAATTCATTAGCAAAATATATATATGTGCAG 144148
QY 210 ----- 210
Db 144149 ACCTGCTAGTACTGGGTAACATCATTTAGTTAAGATTTGGTCTCAGGCCAGCAGGTGG 144208
QY 210 ----- 210
Db 144209 CTCAGCGCTGTAATCCCGACACTTTTGGAGGCCGAGGCTGCAGATCACCTGAGGTGAG 144268
QY 210 ----- 210
Db 144269 AGTTTGAGACGAGCCTGACCAACATGGAGAAACCCCGTGTCTACTAAATAATACAAAATA 144328
QY 210 ----- 210
Db 144329 GCCGAGTGTGGTGGCGCATGCTTATAATGCCAGCTATTTGGAGGCTGAGCAGGAGAAAT 144388
QY 210 ----- 210
Db 144389 CGCTTGAAACCGGGAGGTGGAGGTTGCAGCTGAGATTGCACCATTTGGCTCCAGCC 144448

QY 210 ----- 210
Db 144449 TGGCAACAAGAGACTCCATCAGGAAAAAAGAGATTCTGTCTGAAGGCCACTGC 144508
QY 210 ----- 210
Db 144509 ACCTATTCTCTGTAGGTCTTAGGTTCCCTGTCTAGTTAATTTGGAGGTAATAACCCCTGGCT 144568
QY 210 ----- 210
Db 144569 TACAGGTTGTTTTCAGATTACATAAATAATTTAGGTATAAATTTCTTGGCAGCGTGAAT 144628
QY 210 ----- 210
Db 144629 GGCACATAATGGGTATTTAACTATTAGTTGACTTTTCTGGGTACCTAAAAAGAGAGAGG 144688
QY 210 ----- 210
Db 144689 TTAGGAGATTTTGTCTTGTCTTACTCTCTCTTCCAGACTTTTGTGTAAATACC 144748
QY 210 ----- 210
Db 144749 TCTTACTTCCAGGCTCTTTACTCTCTCTTCCACCCAGATCCTGACGTGTGCATCTCT 144808
QY 211 -----GluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAlaLysL 228
Db 144809 TTGTGCAGGAAATGTGTATCCGAGAAATTTAACTTAAATGAGCTATACAGCGGGCCAAAG 144868
QY 228 ysLeuSerLysAlaGlyAspAsnIleProGluGluInProValAlaSerThrProThr 248
Db 144869 AACTATCAAGGCTGGAGACATATCTCTGAGGACGACCTGTGGCTTCAACCCCAACCA 144928
QY 248 hrValSerAspGlyGluAsnLysLysAspLys 258
Db 144929 CAGTGTCAAGTGGGGAACCAAGAGGATAAA 144960

RESULT 13
AC090218/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-691N7 map 11, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC090218
VERSION
AC090218.3 GI:14718351
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 161397)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A.,
Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehocksky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
```

Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 14, 2001 this sequence replaced gi:13940674:
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12248
Center clone name: 69L_N_7
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 155030 bases at least Q40
Consensus quality: 157856 bases at least Q30
Consensus quality: 159050 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 159797; sum-of-contigs
Quality coverage: 12.0 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
*
* 1 7057: contig of 7057 bp in length
* 7058 7157: gap of 100 bp
* 7158 7519: contig of 362 bp in length
* 7520 7619: gap of 100 bp
* 7620 8839: contig of 1220 bp in length
* 8840 8939: gap of 100 bp
* 8940 9986: contig of 1047 bp in length
* 9987 10086: gap of 100 bp
* 10087 11612: contig of 1526 bp in length
* 11613 11712: gap of 100 bp
* 11713 13287: contig of 1575 bp in length
* 13288 13387: gap of 100 bp
* 13388 14435: contig of 1048 bp in length
* 14436 14535: gap of 100 bp
* 14536 16619: contig of 2084 bp in length
* 16620 16719: gap of 100 bp
* 16720 17857: contig of 1138 bp in length
* 17858 17957: gap of 100 bp
* 17958 19972: contig of 2015 bp in length
* 19973 20072: gap of 100 bp
* 20073 22132: contig of 2060 bp in length
* 22133 22323: gap of 100 bp
* 22323 58899: contig of 36667 bp in length
* 58900 58999: gap of 100 bp
* 59000 70617: contig of 11618 bp in length
* 70618 70717: gap of 100 bp
* 70718 89521: contig of 18804 bp in length
* 89522 89621: gap of 100 bp
* 89622 118253: contig of 28632 bp in length
* 118254 118353: gap of 100 bp
* 118354 155823: contig of 37470 bp in length
* 155824 155923: gap of 100 bp
* 155924 161397: contig of 5474 bp in length.
Location/Qualifiers
1..161397
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source

/chromosome="11"
/map="11"
/clone="RP11-691N7"
/clone_lib="RPC1-11 Human Male BAC"
1..7057
/note="assembly_fragment"
clone_end:SP6
vector_side:left
7158..7519
/note="assembly_fragment"
7620..8839
/note="assembly_fragment"
8940..9986
/note="assembly_fragment"
10087..11612
/note="assembly_fragment"
11713..13287
/note="assembly_fragment"
13388..14435
/note="assembly_fragment"
14536..16619
/note="assembly_fragment"
16720..17857
/note="assembly_fragment"
17958..19972
/note="assembly_fragment"
20073..22132
/note="assembly_fragment"
22233..58899
/note="assembly_fragment"
59000..70617
/note="assembly_fragment"
70718..89521
/note="assembly_fragment"
89622..118253
/note="assembly_fragment"
118354..155823
/note="assembly_fragment"
155924..161397
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 43394 a 34530 c 35670 g 46202 t 1601 others
ORIGIN
Alignment Scores:
Pred. No.: 1.21e-53 Length: 161397
Score: 579.50 Matches: 172
Percent Similarity: 27.37% Conservative: 1
Best Local Similarity: 27.22% Mismatches: 2
Query Match: 42.64% Indels: 459
DB: 2 Gaps: 4
US-09-954-846-2 (1-258) x AC090218 (1-161397)
QY 84 MetPheLeuMetThrCysLysProLeuTyMetGlyProGluTyrlleLysTyPhe 103
Db 38287 GTGTTCTCTGATGACGTGCAACCCCTATATATGGCCCTGAGTATATCAAGTACTTC 38228
QY 104 AsnAspLysThrIleAspGlu----- 110
Db 38227 AATGATAAAACCATGTGATGT-GAGTGTCTCTTTCCCTTCTGTTCCTGGTCCCTGTG 38169
QY 110 ----- 110
Db 38168 GGTGATTTTGTAGTTGTCTCTCCATTCACCTAGGAGGAACAACAGTGCTTCAAAATGGAT 38109
QY 110 ----- 110
Db 38108 GTCACGGGCACCTGTGTTTCATTATGAGAGCCGGGAGAGGAACACAGGATGAGGAAT 38049
QY 110 ----- 110

```
Db 38048 TAGATGCTAAAGTCTGAACCTTCCAGGAGGCTATGTGGATCCAGCTGACATTTTCTTCCC 37989
Qy 111 -----GluLeuGluArgAspLysArgValThrTrpIleValGluPhe 125
Db 37988 TGTATTTGGCAGGAGGAACATAGAACGGGACAAAGAGGTCACTTGGATTTGGAGTTCTTT 37929
Qy 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
Db 37928 GCCAATTGGTCTAATGACTGCCAATCTTTGCCCTATCTATGTCTGACCTCTCCCTTAA- 37870
Qy 145 ----- 145
Db 37869 GTGAGTAGTCAAGAGGAGGATGGTGAATGGAGATGCTGTGCCCTCCCTCTCACTGT 37810
Qy 145 ----- 145
Db 37809 TTTTGGCTTTTCTTTTCTTTTGGCCTTGATTTTCACACATGGTAACCAAGGATCTCC 37750
Qy 145 ----- 145
Db 37749 CTCTCCCTCTTAAATATCTATATCTCCACCTTCTTGATCCATATTTTTCAGCAT 37690
Qy 146 -----TyrAsnCysThrGlyLeuAsnPheGlyLysValAsp 157
Db 37689 ATTAAATAATATATCTTTTCAGATACAACTGTACAGGGCTAAATTTTGGGAAGTGGAT 37630
Qy 158 ValGlyArgTyrThrAspValSerThr----- 166
Db 37629 GTTGGACGCTATACTGATGTTAGTACGGGTATGTAAAGACCTGGGCAGAGGCTGTGAGC 37570
Qy 167 -----Ar 167
Db 37569 AGGGAATCACTTTGAGTGATACATACAGGGACATTTTAGAAGACTTTCTGGGCCCTGCG 37510
Qy 167 gTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGlncl 187
Db 37509 GTACAAAGTGAGCACATCACCCCTCACCAACCACTCCCTACCTGTCTGTTCACAGG 37450
Qy 187 yGlyLysGluAlaMetArgArgProGlnIleAspLysGlyArgAlaValSerTrpTh 207
Db 37449 TGGCAAGGAGGCAATGCGGGGCCACAGATTGACAAGAAAGAGGGGCTGTCTCATGGAC 37390
Qy 207 rPheSerCln----- 210
Db 37389 CTCTCTGAGTACTCTGAAAGGAGGCGAGGTGCATGAAGGTGCAGAACAGTAGTGGG 37330
Qy 210 ----- 210
Db 37329 CTTTCAAGCCCTACCCGGGTTTGATTACAGCTCTGCCACTTGCCCTTACCTTTGAGGG 37270
Qy 210 ----- 210
Db 37269 TGTGGACTAGTTACTAGACCTCATGTTTAAATTTAGCAATATATATTATGTGCAG 37210
Qy 210 ----- 210
Db 37209 ACCTGCTAGTACTGGTAACAATCATTTAGTTAAGATTTGGTCTCAGGCCAGCGAGTGG 37150
Qy 210 ----- 210
Db 37149 CTCAGGCTGTATCCAGCACTTTGGGAGCCGAGGCTGGCAGATCACCTGAGGTGACG 37090
Qy 210 ----- 210
Db 37089 AGTTTGAGACAGCCTGACCAACATGGAGAAACCCCGTGTCTACTTAAATAACAAAAATA 37030
Qy 210 ----- 210
Db 37029 GCCAGTGTGTGGCGCATGCTATATATGCCAGCTATTTGGAGGCTGAGGCAGGAGAA 36970
Qy 210 ----- 210
Db 36969 CGTTGAACCCGGGAGGTGGAGTTGCGAGTGTGAGATTGCACCATTTGCGGTCCAGCC 36910

Qy 210 ----- 210
Db 36909 TGGCAACAAGAGACTCCATCAGGAAAAAAGAGATTCTGTCTGAAGGCCACTGC 36850
Qy 210 ----- 210
Db 36849 ACCTATTGCTCTAGGTGTTAGGTCCCTGTCTCAGTTAATTGGAGGTAAATACCCCTGGCT 36790
Qy 210 ----- 210
Db 36789 TACAGGGTGTGTTTTCAGATTACATTAATAATTTAGGTATAATTTCTTGGCAGCGTGATT 36730
Qy 210 ----- 210
Db 36729 GGCACATAATGGGTATTTAACTATTAGTTGACTTTTCTGGGTACCTAAAAAGAGAAAG 36670
Qy 210 ----- 210
Db 36669 TTAGGAAGATTTTGTCTTGTCTTGTCTTCTCTTCCAGACTTTGTGTAAATACC 36610
Qy 210 ----- 210
Db 36609 TCTTACTTCCAGGCTCTTTTACTCTCCCTTCCAAACCAGACTCTGACGTGTGCATCTCT 36550
Qy 211 -----GluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAlaLysL 228
Db 36549 TTGTGCAGGAGAATGTGATCCGAGAATTTAACTTAAATGAGCTATACCAGCGGCCAAGA 36490
Qy 228 ysLeuSerLysAlaGlyAspAsnIleProGluGlnProValAlaSerThrProThr 248
Db 36489 AACTATCAAGCTGGAGACAATATCCCTGAGGACGCTGTGCTTCAACCCCAACA 36430
Qy 248 hrValSerAspGlyGluAsnLysLysAspLys 258
Db 36429 CAGTGTCAAGTGGGAAAAACAAGAAGATAAA 36398

RESULT 14
AC092707 170540 bp DNA linear HTG 20-JUL-2001
LOCUS Homo sapiens chromosome 11 clone RP11-679G21 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 9 unordered pieces.
ACCESSION AC092707
VERSION AC092707.1 GI:14971453
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170540)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
JOURNAL Homo sapiens chromosome 11, clone RP11-679G21
REFERENCE 2 (bases 1 to 170540)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepli,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrela,S., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,R., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamatares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
```

THEODORE, J., TOPHAM, K., TRAVERS, M., TRAVIS, N., TRIGILLO, J.,
VASSILIEV, H., VIEL, R., VO, A., WILSON, B., WU, X., WYMAN, D., YE, W. J.,
YOUNG, G., ZAINOON, J., ZEMBEK, L., ZIMMER, A. and ZODY, M.
Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12706
Center clone name: 679-G-21
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166430 bases at least Q40
Consensus quality: 168426 bases at least Q30
Consensus quality: 169078 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 169740; sum-of-contigs
Quality coverage: 9.3 in Q20 bases; agarose-fp
Quality coverage: 8.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 9 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 743: contig of 743 bp in length
744 843: gap of 100 bp
844 1842: contig of 999 bp in length
1843 1942: gap of 100 bp
1943 4117: contig of 2175 bp in length
4118 4217: gap of 100 bp
4218 9901: contig of 5684 bp in length
9902 10001: gap of 100 bp
10002 22030: contig of 12029 bp in length
22031 22130: gap of 100 bp
22131 77732: contig of 55602 bp in length
77733 77832: gap of 100 bp
77833 110109: contig of 32277 bp in length
110110 110209: gap of 100 bp
110210 150387: contig of 40178 bp in length
150388 150487: gap of 100 bp
150488 170540: contig of 20053 bp in length.

FEATURES

source
1. 170540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-679G21"
 /clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 743
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
844. 1842
 /note="assembly_fragment"
1943. 4117
 /note="assembly_fragment"
4218. 9901
 /note="assembly_fragment"
10002. 22030

misc_feature
22131. 77732
 /note="assembly_fragment"
misc_feature
77833. 110109
 /note="assembly_fragment"
misc_feature
110210. 150387
 /note="assembly_fragment"
misc_feature
150488. 170540
 /note="assembly_fragment"
 clone_end:17
 vector_side:right
BASE COUNT 47282 a 39323 c 37471 g 45658 t 806 others
ORIGIN
Alignment Scores:
Pred. No.: 1,3e-53 Length: 170540
Score: 579.50 Matches: 172
Percent Similarity: 27.37% Conservative: 1
Best Local Similarity: 27.22% Mismatches: 2
Query Match: 42.64% Indels: 459
DB: 2 Gaps: 4
US-09-954-846-2 (1-258) x AC092707 (1-170540)
QY 84 MetPheLeuMetThrCysLysProLeuTyrMetGlyProGluTyrIleLysTyrPhe 103

Db 61689 GTGTTCTGATGACGTGCAACCCCTATATATGGCCCTGAGTATATCAAGTACTTC 61748
QY 104 AsnAspLysThrIleAspGlu----- 110
 |||||
Db 61749 AATGATAAAACCATTCATGCT-GAGTCTCTTTCCCTTTCTGTTCTTGGGTCCCTGTG 61807
QY 110----- 110
Db 61808 GGTGATTTTGTAGTTGTCTCCTCATTCTAGGAGGAAACAACAGTCTTCAAAATGGAT 61867
QY 110----- 110
Db 61868 GTACGGGCACGTGTGGTTTCATTATGAGAGCGGGGAGAGGAAACACAGGATGAGGAAT 61927
QY 110----- 110
Db 61928 TAGATCTAAAGTCTGAACCTTCCAGAGGCTATGTGGATCCAGCTGACTTTTCTTCCC 61987
QY 111-----GlutLeuGluArgAspLysArgValThrTrpIleValGluPhePhe 125
 |||||
Db 61988 TGTATTTGGCAGGAGGAACTAGAACGGGCAAGAGGTCACCTGGATTGTGGAGTTCTTT 62047
QY 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
 |||||
Db 62048 GCCAATTGGTCTAATGACTGCCCAATCATTTGCCCTATCTATGCTGACCTCTCCCTTAA- 62106
QY 145----- 145
Db 62107 GTGAGTAGTGCAAGGGAGGATGGTGGAAATGGAGATGCTGTGCTTCCCTCTCACTGT 62166
QY 145----- 145
Db 62167 TTTTGGCTTTTCTTTTCTTTTGGCCTTGATTTTACACATGTTACCAAGCAATCTCC 62226
QY 145----- 145
Db 62227 CTCCTCCCTCTTAAATATCTATACTTCCACTTTCTCTTGATCCATTTTTCAGCAT 62286
QY 146-----TyrAsnCysThrGlyLeuAsnProGlyLysValAsp 157
 |||||
Db 62287 ATTAATAATATATTTCTTTTTCAGATACAACTGTACAGGCTAAATTTTGGGAAGGTGAT 62346
QY 158 ValGlyArgTyrThrAspValSerThr----- 166
 |||||
Db 62347 GTTGACGCTATACTGATGTTAGTACGCGGTATGTAAGACCTGGCAGAGGCTCTGACC 62406
QY 167-----Ar 167

* as soon as it is available and the accession number will
* be preserved

* 1 17413: contig of 17413 bp in length
* 17414 17513: gap of 100 bp
* 17514 18244: contig of 731 bp in length
* 18245 18344: gap of 100 bp
* 18345 19084: contig of 740 bp in length
* 19085 19184: gap of 100 bp
* 19185 20416: contig of 1232 bp in length
* 20417 20516: gap of 100 bp
* 20517 21625: contig of 1109 bp in length
* 21626 21725: gap of 100 bp
* 21726 102938: contig of 81213 bp in length
* 102939 103038: gap of 100 bp
* 103039 104905: contig of 1867 bp in length
* 104906 105005: gap of 100 bp
* 105006 108216: contig of 3211 bp in length
* 108217 108316: gap of 100 bp
* 108317 112875: contig of 4559 bp in length
* 112876 112975: gap of 100 bp
* 112976 122676: contig of 9701 bp in length
* 122677 122776: gap of 100 bp
* 122777 138821: contig of 16045 bp in length
* 138822 138921: gap of 100 bp
* 138922 168907: contig of 29986 bp in length
* 168908 169007: gap of 100 bp
* 169008 196861: contig of 27854 bp in length
* 196862 196961: gap of 100 bp
* 196962 199992: contig of 3031 bp in length.

FEATURES

Location/Qualifiers

1..199992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-335H5"
clone_lib="RPC1-11 Human Male BAC"
1..17413
/note="assembly_fragment"
clone_end:SP6
vector_side:left
17514..18244
/note="assembly_fragment"
18345..19084
/note="assembly_fragment"
19185..20416
/note="assembly_fragment"
20517..21625
/note="assembly_fragment"
21726..102938
/note="assembly_fragment"
103039..104905
/note="assembly_fragment"
105006..108216
/note="assembly_fragment"
108317..112875
/note="assembly_fragment"
112976..122676
/note="assembly_fragment"
122777..138821
/note="assembly_fragment"
138922..168907
/note="assembly_fragment"
169008..196861
/note="assembly_fragment"
196962..199992
/note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 54206 a 46381 c 45637 g 52463 t 1305 others
ORIGIN

Alignment Scores:

Pred. No.: 1.59e-53 Length: 19992
Score: 579.50 Matches: 172

Percent Similarity: 27.37%
Best Local Similarity: 27.22%
Query Match: 42.64%
DB: 2 459
Gaps: 4

Conservative: 1
Mismatches: 2

US-09-954-846-2 (1-258) x AC021522 (1-199992)

QY 84 MetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIleLysTyrPhe 103
DB 185240 GTGTTCTTCATGACGTGCAACCCCTTATATATGGCCCTGAGTATATCAAGTACTTC 185181
QY 104 AsnAspLysThrIleAspGlu----- 110
DB 185180 AATGATAAAACCATTTGATGT-CAGTGTCTCTTTCCCTTTCTGTCTTGGTCCCTTGTG 185122
QY 110 ----- 110
DB 185121 GGTGATTTGTAGTTGTCTCTCCATTCAC TAGGAGGAACAACAGTGCTTCAAAATGGAT 185062
QY 110 ----- 110
DB 185061 GTCACGGCCTGTGTTTCAATTATGAGAGCCGGGAGAGGAACACAGGATGAGGAAT 185002
QY 110 ----- 110
DB 185001 TAGATGCTAAAGCTCTGAACCTTCCAGGAGGCTATGTGGATCCAGCTGACTTTTCTTCCC 184942
QY 111 ----- GluLeuGluArgAspLysArgValThrTriPileValGluPhePhe 125
DB 184941 TGTATTTGGCAGGAGGAACACTAGACGGGACAGAGGCTCATTGGATGCTGGAGTCTTT 184882
QY 126 AlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAlaAspLeuSerLeuLys 145
DB 184881 GCCAATGTGCTTAATGACTGCTCAATCTTGGCCCTATCTATGCTGACCTCTCCCTTAA- 184823
QY 145 ----- 145
DB 184822 GTGAGTAGTGCAAGGAGGGGATGTGGAAATGGAGATGCTGTGCCTTCCCTCTCACTGT 184763
QY 145 ----- 145
DB 184762 TTTTGGCTTTCTTTTCTTTTGGCCTTGATTTTTCACACATGGTAACCAAGCATCTCC 184703
QY 145 ----- 145
DB 184702 CTCCTCCCTCTAAATATATCTATCTTCCACTTTCCTTGATCCATTATTTTTCAGCAT 184643
QY 146 ----- TyrAsnCysThrGlyLeuAsnPheGlyLysValAsp 157
DB 184642 ATTAATAATATATTTCTTTTCAGATACAACTGTACAGGCTAAATTTTGGGAAGGTGAT 184583
QY 158 ValGlyArgTyrThrAspValSerThr----- 166
DB 184582 GTTGGACCTATACTGATGTTAGTACGCGGTATGTAAGACCTGGGACAGGGCTTGACC 184523
QY 167 ----- Ar 167
DB 184522 AGGGAATCACTTTGAGTGATACATACAGGGACATTTAGAGAACTTTCTGGGCCCTGCAG 184463
QY 167 gTyrLysValSerThrSerProLeuThrLysGlnLeuProThrLeuIleLeuPheGlnG1 187
DB 184462 GTACAAAGTGAGCACATCACCCCTCACCAGCAACTCCCTACCTGTATCTCTTCCAGG 184403
QY 187 yGlyLysGluAlaMetArgArgProGlnIleAspLysLysGlyArgAlaValSerTriPh 207
DB 184402 TGGCAAGGAGCAATGCGGGGCCACAGATTGACAAGAAAGACGGGCTGTCTCATGGAC 184343
QY 207 rPheSerGlu----- 210
DB 184342 CTTCTCTGAGGTACCTGAAGGAGGGCAGGTGCATGAAGGTGCAGACACAGTAGTGGG 184283
QY 210 ----- 210

```
Db 184282 CTTTCAAGCCCTACCCGGGTTTGATTACAGCTCTGCCACTTGCCTCCCATTTAGCTTTGAGGG 184223
QY 210 ----- 210
Db 184222 TGTGGACTAGTTACTAGACCTCATGTTTAAATTCATTAGCAAAATATATATTATGTGCAG 184163
QY 210 ----- 210
Db 184162 ACCTGCTAGTACTGGGTACAAATCATTTAGTTAGATTGGTCTCAGGCCAGGCACGGTGG 184103
QY 210 ----- 210
Db 184102 CTCACGCTGTAAATCCAGCACCTTTGGGAGCGGAGGTGGCAGATCACCTGAGGTCAGG 184043
QY 210 ----- 210
Db 184042 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCGGTGTCTACTAAAAATACAAAAATA 183983
QY 210 ----- 210
Db 183982 GCCGAGTGTGGTGGCCCATGCCCTATAATGCCAGCTATTTGGGAGGCTGAGGCAGGAGAAT 183923
QY 210 ----- 210
Db 183922 CGCTTGAACCCGGGAGGTGGAGGTTCAGTGAGCTGAGATTGCACCATTCGCGTCCAGCC 183863
QY 210 ----- 210
Db 183862 TGGGCAACAAGAGACTCCATCAGGAAAAAAGAAATTCGTCTCGAAGGCCACTGC 183803
QY 210 ----- 210
Db 183802 ACCTATTGCTGTAGTGTAGTTCCTCCCTGTCAGTTAATTTGGAGGTAATAACCCCTGGCT 183743
QY 210 ----- 210
Db 183742 TACAGGGTTGTTTGCAGATTACATAAATAATTTAGGTATATATCTTGGCACGGTGATT 183683
QY 210 ----- 210
Db 183682 GGCACATAATGGGTATTTAACTATTAGTTGACTTTTCTGGGTACCTAAAAAGAGGAAGG 183623
QY 210 ----- 210
Db 183622 TTAGGAAGATTTTGTCTCTTGTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 183563
QY 210 ----- 210
Db 183562 TCCTTACTCCAGGCTCTTTACTCTCCCTTCCAAACCCAGATCCTGAGGTGTCATCTCTT. 183503
QY 211 -----GluAsnValIleArgGluPheAsnLeuAsnGluLeuTyrGlnArgAlaLysL 228
|||||
183502 TTGTGCAGGAGAATGTGATCCGAGAATTTAACTTAAATGAGCTATACCAAGCGGCCAAGA 183443
QY 228 ysLeuSerLysAlaGlyAspAsnIleProGluGluGlnProValAlaSerThrProThrT 248
|||||
183442 AACTATCAAGGCTGGAGCAATATCCCTGAGGAGCAGCCTGTGGCTTCAACCCCCACCA 183383
QY 248 hrValSerAspGlyGluAsnLysLysAspLys 258
|||||
Db 183382 CAGTGTACATGGGGAACACAGAGATAAA 183351
```

Search completed: July 12, 2003, 11:49:39
Job time : 2172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2003, 09:18:57 ; Search time 1606 Seconds
(without alignments)
15681.184 Million cell updates/sec

Title: US-09-954-846-4
Perfect score: 1555
Sequence: 1 AGGGGAGCGGGCGGAGACC.....TAATAAAGATTGGGATTA 1555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

--arched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estim.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_estl.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_eston.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	950.8	61.1	980	9	AL515716
c 2	869	55.9	1014	9	AL563867
3	866.4	55.7	870	14	BQ687961
4	855.8	55.0	1022	13	BM555220
5	835	53.7	919	14	BQ961752
6	826.2	53.1	911	14	BQ940063

c 7	823.8	53.0	832	9	AL571100	AL571100
8	801.2	51.5	998	14	BM915245	BM915245
9	796.2	51.2	884	14	BQ430136	BQ430136
c 10	776	49.9	1017	13	B1524823	B1524823
11	765.8	49.2	838	14	BQ690091	BQ690091
12	763.2	49.1	835	14	BQ962329	BQ962329
13	751.6	48.3	913	14	BQ920164	BQ920164
14	750.8	48.3	894	14	BQ687916	BQ687916
c 15	746.2	48.0	786	9	AL569941	AL569941
16	741	47.7	1009	12	BF526515	BF526515
17	737.4	46.8	793	13	B1771888	B1771888
c 18	728.4	46.8	767	14	BM974314	BM974314
19	716.8	46.1	911	12	BG762562	BG762562
20	715.8	46.0	1097	14	BQ073145	BQ073145
21	713.8	45.9	827	12	BE908352	BE908352
c 22	712.2	45.8	968	12	BE748905	BE748905
23	704.4	45.3	918	14	BQ945418	BQ945418
c 24	694.2	44.6	1052	12	BF569320	BF569320
25	691.6	44.5	858	14	BQ686696	BQ686696
c 26	689.4	44.3	720	14	BM974059	BM974059
27	688	44.2	724	12	BE741788	BE741788
28	686.6	44.2	1008	12	BE898497	BE898497
c 29	684.4	44.0	730	14	BM988558	BM988558
30	684	44.0	690	14	BM762654	BM762654
c 31	683.4	43.9	740	14	BM975449	BM975449
32	682	43.9	834	12	BG179545	BG179545
c 33	681.2	43.8	702	9	AL570762	AL570762
34	681.2	43.8	948	12	BE796785	BE796785
35	679.8	43.7	880	10	BE616524	BE616524
36	679	43.7	860	13	B1161293	B1161293
37	678.6	43.6	928	12	BG251786	BG251786
38	677.8	43.6	1142	13	BM559140	BM559140
c 39	676.6	43.5	710	14	BQ447137	BQ447137
40	673.6	43.3	766	12	BG762793	BG762793
c 41	671.4	43.2	698	14	BQ009303	BQ009303
c 42	670.4	43.1	731	14	BM975356	BM975356
c 43	669.6	43.1	779	14	BM981327	BM981327
44	669.4	43.0	838	12	BF342782	BF342782
c 45	662.2	42.6	716	14	BM978725	BM978725

ALIGNMENTS

RESULT 1
AL515716/c
LOCUS AL515716 980 bp mRNA linear EST 13-FEB-2001
DEFINITION AL515716 LTI_NFL011_NBC1 Homo sapiens CDNA clone CSODA001YG10.3
ACCESSION AL515716
VERSION AL515716.1 GI:12779209
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 980)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..980

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODA001YG10"
/clone_lib="LTI_NFL011_NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"

/note="Organ: brain; Vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 264 a 229 c 229 g 253 t 5 others
ORIGIN

Query Match 61.1%; Score 950.8; DB 9; Length 980;
Best Local Similarity 98.0%; Pred. No. 2.4e-286;
Matches 953; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 554 ACACGCTACAAAGTACGACATCACCCCTCACCAAGCACTCCCTACCTGATCCTGTC 613
DB 974 ACACGCTACAAAGTACGACATCACCCCTCACCAAGCACTCCCTACCTGATCCTGTC 915
614 CAAAGTGGCAAGGAGGCAATGCGCGGCCACAGATTGACAAAGGAGCGGCTGCTCA 673
914 CAAAGTGGCAAGGAGGCAATGCGCGGCCACAGATTGACAAAGGAGCGGCTGCTCA 855
QY 674 TGAACCTCTCTGAGGAGATGTATCCGAGATTTAACTTAATGAGCTATACGAGG 733
DB 854 TGAACCTCTCTGAGGAGATGTATCCGAGATTTAACTTAATGAGCTATACGAGG 795
QY 734 GCCAAGAAATATCAAGGCTGAGACAATATCCTCAGGAGCAGCTGTGGCTTCAAC 793
DB 794 GCCAAGAAATATCAAGGCTGAGACAATATCCTCAGGAGCAGCTGTGGCTTCAAC 735
QY 794 CCACACACAGTGCAGATGGGAAACAAAGAGGATTAATGATCTCTCACTTTGGCAGT 853
DB 734 CCACACACAGTGCAGATGGGAAACAAAGAGGATTAATGATCTCTCACTTTGGCAGT 675
QY 854 GCTTCCTCTCTGATCAATCCAGCTCTTCCATTAACCAAGCCTGAGGCTGAGGCTT 913
DB 674 GCTTCCTCTCTGATCAATCCAGCTCTTCCATTAACCAAGCCTGAGGCTGAGGCTT 615
QY 914 TTATTATGTTTCCCTTTGGCTGTGACTGGTGGGCGCAGCATCAGCTCTGATTTAA 973
DB 614 TTATTATGTTTCCCTTTGGCTGTGACTGGTGGGCGCAGCATCAGCTCTGATTTAA 555
QY 974 AGAGGCATCTAGGGAATTTGACGACCCCTACAGGAAGGCTGCGATGCTGTGGCCAACT 1033
DB 554 AGAGGCATCTAGGGAATTTGACGACCCCTACAGGAAGGCTGCGATGCTGTGGCCAACT 495
QY 1034 GTTTCAGTGAAGCAAGAGATCTCATAGGCGGAGGGGAAATGTTTCCCTCCAAAG 1093
DB 494 GTTTCAGTGAAGCAAGAGATCTCATAGGCGGAGGGGAAATGTTTCCCTCCAAAG 435
QY 1094 CTTGGGTGAGTGTAACTGCTTATCAGCTATTTCAGCATCTCCATGTTTCCCATGA 1153
DB 434 CTTGGGTGAGTGTAACTGCTTATCAGCTATTTCAGCATCTCCATGTTTCCCATGA 375
QY 1154 AACTCTGTGTTTCATCATCTCTTCTAGTGTGACCTGCACAGCTTGGTTAGACCTAGATT 1213
DB 374 AACTCTGTGTTTCATCATCTCTTCTAGTGTGACCTGCACAGCTTGGTTAGACCTAGATT 315
QY 1214 TAACCTTAAGGTAGATGCTGGGGTATAGAACGCTTAAGAAATTTCCCCCAAGGACTCTTG 1273
DB 314 TAACCTTAAGGTAGATGCTGGGGTATAGAACGCTTAAGAAATTTCCCCCAAGGACTCTTG 255
QY 1274 CTTCCCTTAAGCCCTCTGCTTCTGTTATGCTTCAATTAAGCTTAAGCTTAAGCTTGTG 1333
DB 254 CTTCCCTTAAGCCCTCTGCTTCTGTTATGCTTCAATTAAGCTTAAGCTTGTGTTATG 195
QY 1334 TCCTAGTCTTAAGGAGAAACCTTTTAACCAACAAAGTTTATCATTTGAAGCAATATTGA 1393
DB 194 TCCTAGTCTTAAGGAGAAACCTTTTAACCAACAAAGTTTATCATTTGAAGCAATATTGA 135

QY 1394 ACACCCCTATTGTTGGGATTGAGAGGGTGAATAGAGGCTTGAGACTTTCCCTTTG 1453
DB 134 ACACCCCTATTGTTGGGATTGAGAGGGTGAATAGAGGCTTGAGACTTTCCCTTTG 75
QY 1454 TGTGGTAGGACTTGGAGGAGAAATCCCTGACCTTCTACTAACCTCTGACATACCTCC 1513
DB 74 TGTGGTAGGACTTGGAGGAGAAATCCCTGACCTTCTACTAACCTCTGACATACCTCC 15
QY 1514 ACAC 1518
DB 14 GCAVC 10

RESULT 2
AL563867/c
LOCUS
DEFINITION AL563867 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD008YG11 3
prime, mRNA sequence.
ACCESSION AL563867
VERSION AL563867.1 GI:12913683
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1014)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 1014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD008YG11"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 277 a 230 c 229 g 248 t 30 others
ORIGIN

Query Match 55.9%; Score 869; DB 9; Length 1014;
Best Local Similarity 91.4%; Pred. No. 1.1e-260;
Matches 909; Conservative 24; Mismatches 60; Indels 2; Gaps 2;

QY 486 CCCTTAATCAACTGTACAGGCTAAATTTTGGGAAGGTGATGTTGGAGCTTACTG 545
DB 1007 CCTTAAAAACACTGAACAGGCTAAATTTGGGAAGGTGATGTTGGAGCTTAACTG 948
QY 546 ATGTTAGTAGCGGTGACAAAGTGAACATCACCCCTCACCAAGCAACTCCCTACCTGA 605
DB 947 ATGTTGAACGCGGTACAAAGGAGACWTACCCCTCMCCARGCAACTCCCTACCTGA 888
QY 606 TCCGTGTTCCAAAGTGGCAAGGAGCAATGCGGCGGCCACAGATTGACAAGAAAGACGGG 665
DB 887 TCCTGTCCCAAGTGGCAAGGAGCAATGCGGCGGCCACCAATTAACAAGAAAGACGGG 828
QY 666 CTGTCTCATGACCTCTCTCTGAGGAGATGTGATCCGAGAAATTTAACTTAATGAGCTAT 725

```
Db 827 YTTTTCATGGACCTCTCTGAGGARAATKTGATCCGAGAATTTAACTTAAATGAGCTAT 768
Qy 726 ACCAGCGGGCAAGAACTATCAAGAGCTGGAGACAATATCCCTGAGAGAGCCTGTGG 785
Db 767 ACCGCGGGCAAGAACTATCAAGAGCTGGAGACAATATCCCTGAGGCGACGTGTGG 708
Qy 786 CTTCAACCCCCACACACAGTGTGAGATGGGAAAACAAGAGGATAAATAAGATCTCTACT 845
Db 707 CTTCAACCCCCACACACAKTKCAGATGGGAAWACAGAAGATATAAGACCTTMACT 648
Qy 846 TTGGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 905
Db 647 TTGGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 588
Qy 906 GCAGCCTTTTATTTATTTTCCCTTTTCCCTTTTCCCTTTTCCCTTTTCCCTTTTCCCT 965
Db 587 GCAGCCTTTTATTTATTTTCCCTTTTCCCTTTTCCCTTTTCCCTTTTCCCTTTTCCCT 528
Qy 966 GATTTTAAAGAGGATCTAGGGAATTTGTAGGACCCCTACAGGAAGGCTGCCATGCTGT 1025
Db 527 GATTTTAAAGAGGATCTAGGGAATTTGTAGGACCCCTACAGGAAGGCTGCCATGCTGT 468
Qy 1026 GGCCAACTGTTTCTACTGGAGCAAGAAAGAGATCTATAGGACGGGAGGAAATGTTTC 1085
Db 467 GGCCAACTGTTTCTACTGGAGCAAGAAAGGCTCTCATAGGACGGGAGGAAATGTTTC 408
Qy 1086 CCTCAAGCTGGGTGAGTGTAACTGCTTATCAGCTATTCAGATATTCAGATATTCAGAT 1145
Db 407 CCTCAAGCTGGGTGAGTGTAACTGCTTATCAGCTATTCAGATATTCAGATATTCAGAT 349
Qy 1146 CTCCATGAACCTGTGTTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1205
Db 348 CTCCATGAACCTGTGTTTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 289
Qy 1206 CCTAGATTTAACCTAAGGTAAAGTGTCTGAGGATATAGAACGCTAAGAAATTTTCCCAAG 1265
Db 288 CCTAGATTTAACCTAAGGTAAAGGCTGAGGATATAGAACGCTAAGAAATTTTCCCAAG 229
Qy 1266 GACTCTGCTCTCTTAAAGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1325
Db 228 AACTCTGCTCTCTTAAAGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 169
Qy 1326 TAACCTTGTGCTGCTTAAAGCAAACTTTTAAACCAAAAGTTTATCATTAAGAAC 1385
Db 168 TAACCTTGTGCTGCTTAAAGCAAACTTTTAAACCAAAAGTTTATCATTAAGAAC 109
Qy 1386 AATATTGAACACCCCTCTTTTCTGGGATGAGAGGGGTGATAGAGGCTTTGAGACT 1445
Db 108 AATATTGAACACCCCTCTTTTCTGGGATGAGAGGGGTGATAGAGGCTTTGAGACT 49
Qy 1445 TTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480
Db 48 TTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15
```

```
RESULT 3
B0687961
LOCUS
DEFINITION B0687961 870 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
ACCESSION B0687961
VERSION B0687961.1 GI:21813277
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 870)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
```

Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2388 row: 0 column: 05
High quality sequence stop: 647.

FEATURES

Location/Qualifiers
1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6248668"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pMTA7; Site:1: XhoI;
Site:2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT

227 a 203 c 215 g 224 t

ORIGIN

Query Match 55.7%; Score 866.4; DB 14; Length 870;
Best Local Similarity 99.8%; Pred. No. 6.8e-260;
Matches 867; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 369 ATGATAAAACCATTTGATGAGGAACATAGAACGGGACAGAGGCTCACTTGGATTGGAGCT 428
Db 1 ATGATAAAACCATTTGATGAGGAACATAGAACGGGACAGAGGCTCACTTGGATTGGAGCT 60
Qy 429 TCTTTGCCAATTTGGTCTAATGACTGCCAATCATTTTGGCCCTATCTATGCTGACCTCTCC 488
Db 61 TCTTTGCCAATTTGGTCTAATGACTGCCAATCATTTTGGCCCTATCTATGCTGACCTCTCC 120
Qy 489 TTAATAACAACCTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGACCTATACCTATG 548
Db 121 TTAATAACAACCTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGACCTATACCTATG 180
Qy 549 TTAGTACCGGGTACAAAGTGAGCAGCATCACCCCTCACCAAGCAACTCCCTACCCCTGATCC 608
Db 181 TTAGTACCGGGTACAAAGTGAGCAGCATCACCCCTCACCAAGCAACTCCCTACCCCTGATCC 240
Qy 609 TGTTCGAAGTGGCAAGGAGGCAATGCGGGGCCACAGATTGACAAGAAAGGACGGGTG 668
Db 241 TGTTCGAAGTGGCAAGGAGGCAATGCGGGGCCACAGATTGACAAGAAAGGACGGGTG 300
Qy 669 TCTCATGAGCCTCTCTCTGAGGAGATGTATCCCGAGAATTTAACTTAATGAGCTATACC 728
Db 301 TCTCATGAGCCTCTCTCTGAGGAGATGTATCCCGAGAATTTAACTTAATGAGCTATACC 360
Qy 729 AGCGGGCCCAAGAACTATCAAAAGGCTGGAGACAATATCCCTGAGGACGAGCCTGTGGCTT 788
Db 361 AGCGGGCCCAAGAACTATCAAAAGGCTGGAGACAATATCCCTGAGGACGAGCCTGTGGCTT 420
Qy 789 CAACCCCCACACAGTGTCTGATGGGAAAACAAGAGGATAAATAAGATCTCTCACTTTG 848
Db 421 CAACCCCCACACAGTGTCTGATGGGAAAACAAGAGGATAAATAAGATCTCTCACTTTG 480
Qy 849 GCAGTGTCTCTCTCTCTGCTCAATTCAGGCTCTTTCCATACCACCAAGCCTGAGGCTGCA 908
Db 481 GCAGTGTCTCTCTCTCTGCTCAATTCAGGCTCTTTCCATACCACCAAGCCTGAGGCTGCA 540
Qy 909 GCCTTTTATTTATGTTTTCCTTTTGGCTGTGACCTGGGTGGGCGAGCAGCTGAGCTTCTGAT 968
Db 541 GCCTTTTATTTATGTTTTCCTTTTGGCTGTGACCTGGGTGGGCGAGCAGCTGAGCTTCTGAT 600

969	QY	TTTTAAAGAGCATCTTAGGGAATTTGTCAGGCACCCCTACAGGAAGCCCTGCCATGCTGTGGC	1028
601	Db	TTTTAAAGAGGCATCTTAGGGAATTTGTCAGGCACCCCTACAGGAAGCCCTGCCATGCTGTGGC	660
1029	QY	CAACTGTTTTCACATGGAGCAAGAAGAGATCTCATAGGACGGAGGGGAAATGGTTTCCCT	1088
661	Db	CAACTGTTTTCACATGGAGCAAGAAGAGATCTCATAGGACGGAGGGGAAATGGTTTCCCT	720
1089	QY	CCAAGCTGGGTTCAGTGTGTTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTC	1148
721	Db	CCAAGCTGGGTTCAGTGTGTTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTC	780
1149	QY	CATGAACACTCTGGGTTTTCATCATTCCTTCTTAGTTCAGCTGCACAGCTTGGTTAGACCT	1208
781	Db	CATGAACACTCTGGGTTTTCATCATTCCTTCTTAGTTCAGCTGCACAGCTTGGTTAGACCT	840
1209	QY	AGATTTAAACCTTAAGGTAAGATCTGGGG	1237
841	Db	AGATTTAACCTTAAGGTAAGATCTGGGG	869
RESULT 4			
- "55220			
JS			
- "INITIATION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY			
Db			
QY			
122			

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2567 row: h column: 13
High quality sequence stop: 727.
Location/Qualifiers
1. .919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6379836"
/clone_lib="NIH_MGC_42"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
223 a 238 c 222 g 228 t 8 others

BASE COUNT 223 a 238 c 222 g 228 t 8 others
ORIGIN

Query Match 53.7%; Score 835; DB 14; Length 919;
Best Local Similarity 97.8%; Pred. No. 4.9e-250;
Matches 874; Conservative 0; Mismatches 17; Indels 3; Gaps 3;

QY 49 GCGCGAAGATGCGGCTTGGACCTCTTAATGCTCTCGTATTCGGTGGCGCGACT 108
|||
Db 13 GGGCGGAAGATGCGGCTTGGACCTCTTAATGCTCTCGTATTCGGTGGCGCGACT 72
|||
QY 109 TTACGATGGCTGCCCAACCTTACTTCTCTCGGCGCTCTCTCTGCTGCTTCTCT 168
|||
Db 73 TTACGATGGCTGCCCAACCTTACTTCTCTCGGCGCTCTCTCTGCTGCTTCTCT 132
|||
QY 169 ACTCGTGAGAACTGCGCGCTCTGCCAGGCTCTGCCACCCAGCGAAGAGCTAA 228
|||
Db 133 ACTCGTGAGAACTGCGCGCTCTGCCAGGCTCTGCCACCCAGCGAAGAGCTAA 192
|||
QY 229 CCGGTGTGACTTGTGAGAGAGAGTGGAGATCCTGATGTTCTCAGTGCCATTGTAT 288
|||
Db 193 CCGGTGTGACTTGTGAGAGAGAGTGGAGATCCTGATGTTCTCAGTGCCATTGTAT 252
|||
QY 289 GATGAAGAACCGAGATCCATGTCCTGATGACGTGCAACCCCTTATATATATATAT 348
|||
Db 253 GATGAAGAACCGAGATCCATGTCCTGATGACGTGCAACCCCTTATATATATATAT 312
|||
QY 349 TGAGTATATCAAGTACTTCAATGATTAACCAATGATGAGGAAGTACAGCGGACAAG 408
|||
Db 313 TGAGTATATCAAGTACTTCAATGATTAACCAATGATGAGGAAGTACAGCGGACAAG 372
|||
QY 409 GTGACTTGGATTGTGAGTCTTTTGCCAAATGCTTAATGATGACGCCCAATATTTGCC 468
|||
Db 373 GTGACTTGGATTGTGAGTCTTTTGCCAAATGCTTAATGATGACGCCCAATATTTGCC 432
|||
QY 469 TATCTATGCTGACCTCTCCCTTAATCAACTGATACAGGCTTAATTTTGGAGAGTGA 528
|||
Db 433 TATCTATGCTGACCTCTCCCTTAATCAACTGATACAGGCTTAATTTTGGAGAGTGA 492
|||
QY 529 TGTGAGCGCTATGATGATTTAGTACGCGGTACAAAGTACATACCCCTCACCAC 588
|||
Db 493 TGTGAGCGCTATGATGATTTAGTACGCGGTACAAAGTACATACCCCTCACCAC 552
|||
QY 589 GCAACTCCCTTACCTGATCTGTTCCAAAGTGGCAAGGAGGCAATGCGGCGGCACAGAT 648
|||
Db 553 GCAACTCCCTTACCTGATCTGTTCCAAAGTGGCAAGGAGGCAATGCGGCGGCACAGAT 612
|||

649 TGACAAGAAAGGACGGGCTGCTCATGGACCTTCTCTGAGGAGAATGTATCCGGAAT 708
|||||
613 TGACAAGAAAGGACGGGCTGCTCATGGACCTTCTCTGAGGAGAATGTATCCGGAAT 672
|||||
709 TAACTTAAATGAGCTATACCAAGGCGGCAAGAACTATCAAAAGGCTGGAGACAATATCCC 768
|||||
673 TAACTTAAATGAGCTATACCAAGGCGGCAAGAACTATCAAAAGGCTGGAGACAATATCCC 732
|||||
769 TGAGGAGACGCTGTGGCTTCAACCCCAACCCACAGAGTGTGAGATGGGGAACAGAGGA 828
|||||
733 TGAGGAGACGCTGTGGCTTCAACCCCAACCCACAGAGTGTGAGATGGGGAACAGAGGA 792
|||||
829 TAAATAAGATCTCACTTTGGCAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 887
|||||
793 TAAATAAGATCTCACTTTGGCAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 852
|||||
888 AACCAAC-AAGCCTGAGGCTGACGCTTTT-ATTATGTTTCTCTCTCTCTCTCTCTCTCTCT 939
|||||
853 TACCACAAACCTGAGGCTGACGCTTTTAAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 906
|||||

RESULT 6
BQ940063
LOCUS
DEFINITION
AGENCOURT_8865577 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6341656
5', mRNA sequence.
ACCESSION BQ940063
VERSION BQ940063.1 GI:22355541
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 911)
AUTHORS NIH-MGC http://mgi.mcg.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2540 row: a column: 17
High quality sequence stop: 727.
Location/Qualifiers
1. .911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6341656"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
236 a 206 c 223 g 245 t 1 others

BASE COUNT 236 a 206 c 223 g 245 t 1 others
ORIGIN

Query Match 53.1%; Score 826.2; DB 14; Length 911;
Best Local Similarity 97.0%; Pred. No. 2.8e-247;
Matches 875; Conservative 0; Mismatches 19; Indels 8; Gaps 3;

588	QY	AGCAACTCCCTACCTGATCTGTTTCAAGTGTGCAGGAGGCAATCGCGCGGCACAGA	647
1	Db	AGCAACTCCCTACCTGATCTGTTTCAAGTGTGCAGGAGGCAATCGCGCGGCACAGA	60
648	QY	TTCCACAAGAAGAGCAGGCGTCTCTCATGAGCCTTCTCTGAGGAGAAATGTGATCCGAGAAT	707
61	Db	TTCCACAAGAAGAGCAGGCGTCTCTCATGAGCCTTCTCTGAGGAGAAATGTGATCCGAGAAT	120
708	QY	TTTAACCTTAATGAGCTATACACGCGGCCCAAGAAACTATCAAAAGCGTGGAGACAAATATCC	767
121	Db	TTTAACCTTAATGAGCTATACACGCGGCCCAAGAAACTATCAAAAGCGTGGAGACAAATATCC	180
768	QY	CTGAGGAGCAGCCTGTGGCTTTCAACCCCCACACAGTGTGAGATGGGGAAAAACAAGAAGG	827
181	Db	CTGAGGAGCAGCCTGTGGCTTTCAACCCCCACACAGTGTGAGATGGGGAAAAACAAGAAGG	240
828	QY	ATAAATAAGATCCTCACTTTGGCAGTCTTCCCTCTCCTCTCAATCCAGGCTCTTTCCAT	887
241	Db	ATAAATAAGATCCTCACTTTGGCAGTCTTCCCTCTCCTCTCAATCCAGGCTCTTTCCAT	300
888	QY	AACCACAAGCCTGAGGCTGCAGCCTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTG	947
301	Db	AACCACAAGCCTGAGGCTGCAGCCTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTG	360
948	QY	GGCAGCATGCAGTCTCTGATTTTAAAGAGGCATCTAGGGAATGTGTCAGGCACCCCTACAG	1007
361	Db	GGCAGCATGCAGTCTCTGATTTTAAAGAGGCATCTAGGGAATGTGTCAGGCACCCCTACAG	420
1008	QY	GAAGGCTGCCATCTCTGTGCCAACTCTTTCACCTGGAGCAAGAAGATCTCATAGGAC	1067
421	Db	GAAGGCTGCCATCTCTGTGCCAACTCTTTCACCTGGAGCAAGAAGATCTCATAGGAC	480
1068	QY	GGAGGGGGAATGGTTTCCCTCCCAAGCTTGGGTCAGTGTGTTAACTGCTTATCAGCTATT	1127
481	Db	GGAGGGGGAATGGTTTCCCTCCCAAGCTTGGGTCAGTGTGTTAACTGCTTATCAGCTATT	540
1128	QY	CAGACATCTCCATGGTTTCTCCATGAAACTCTGTGGTTTCATCATCTCTTAGTTGAC	1187
541	Db	CAGACATCTCCATGGTTTCTCCATGAAACTCTGTGGTTTCATCATCTCTTAGTTGAC	600
1188	QY	CTGCACAGCTTGGTTAGACCTAGATTTAAACCCTAAGGTAAGATGCTGGGGTATAGAAGC	1247
601	Db	CTGCACAGCTTGGTTAGACCTAGATTTAAACCCTAAGGTAAGATGCTGGGGTATAGAAGC	660
1248	QY	TAAGAAATTTTCCCCCAAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTATGGTCT	1307
661	Db	TAAGAAATTTTCCCCCAAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTATGGTCT	720
1308	QY	TCATTAAGAAGTATAAGCCTAACCTTTGTCGTAGTCTTAAGGAGAAACCTTTAACCCACAA	1367
721	Db	TCATTAAGAAGTATAAGCCTAACCTTTGTCGTAGTCTTAAGGAGAAACCTTTAACCCACAA	780
1368	QY	GTTTTATCATTTGAAGACAATATTGAACAACCCCTCTTATTTGTGGGATTCGAGAAGG	1424
781	Db	GTTTTATCATTTGAAGACAATATTGAACAACCCCTCTTATTTGTGGGATTCGAGAAGG	840
1425	QY	GGTGAATPAGA--GGCTTTGAGACTTTCC---TTTGTGGTGGTAGGACTTTGGAGGAGAAATCC	1479
841	Db	GGTGAATPAGGGCTTGAGACTTTTCCCTTTTGGTGTGGTAGGACTTTGGGAGGGAGAAAT	900
1480	QY	CC 1481	
901	Db	CC 902	
RESULT 7			
AL571100/c			
LOCUS			
DEFINITION			
AL571100 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI027YB05 3			
prime, mRNA sequence.			
ACCESSION			
AL571100			
VERSION			
AL571100.1 GI:12928060			
KEYWORDS			
EST.			

	Query Match	49.9%	Score 776;	DB 13;	Length 1017;
	Best Local Similarity	91.1%	pred. No. 1.6e-231;		
	Matches 927;	Conservative 0;	Mismatches 75;	Indels 16;	Gaps 9;
Qy	507	GGCTAAATTTGGGAAGGTGGATGTTGGACGCTATCTACTGATGTTAGTACCGGTTACAAG	566		
Db	1016	GGCCAAACTTCGGAAACGCTGATTTGGGACCAAACTACGGTAAAGCCGCGGTACAAG	957		
Qy	567	TGAGCATACACCCCTCACCAAGCAAG--TCCCTACCCCTGATCCTGTTTCCAA--GGTGG	621		

```

Db      956 GGGAGTAATACCCCTTCCCAAGCGAGGCTTCCTTACCGGACCCCGTGTTCCTCAACGGTGGC 897
Qy      622 CAAGGAGGCAATCGCGCGGCCA-CAGATTGACAAGAAAGGAC- ---GGGCTGTCTCATGGA 677
Db      896 AAGGAGCGCAATCGCGCGGCCAGCAGATGACAGAAAGAGCCGGCTGGTCTCAGGGA 837
Qy      678 CTTCTCTGAGGAGATGTATCCGAGAAATTAACATTAATAGCTATACCAAGCGGGCCA 737
Db      836 CTTCTCTGAGAGAAATGTATCCGAGAAATTAACATTAATAGCTATACCAAGCGGGCCA 777
Qy      738 AGAACTATCAAGGCTGGAGACAATATCCCTGA-CGAGCAGCCTGTGCTTCAACCCCC 796
Db      776 GAACTTATCAAGGCTGGAGACAATATCCGAGGAGCAGCCTGTGCTTCAACCCCC 717
Qy      797 ACCACAGTGTAGATCGGGAAACAGAAAGGATAAATAAGATCTCTCACCTTTGGCAGTGCT 856
Db      716 ACCACAGTGTAGATCGGGAAACAGAAAGGATAAATAAGATCTCTCACCTTTGGCAGTGCT 657
Qy      857 TCCTCTCTGCTCAATTCAGGCTCTTTTCCAT-AACACAAAGCCTGAGGCTGCAGCCTTTT 915
Db      656 TCCTCTCTGCTCAATTCAGGCTCTTTTCCATGAACACACAGCCTGAGGCTGCAGCCTTGT 597
Qy      916 ATTTATGTTTCCCTTGGCTGTGACTGGTGGGCGAGCATGCAGCTTCTGATTTAAAG 975
Db      596 ATTTATGTTTCCCTTGGCTGTGACTGGTGGGCGAGCATGCAGCTTCTGATTTAAAG 537
Qy      976 AGGCATCTAGGAAATGTGAGCACCCTACAGAAAGCCTGCATGCTGTGGCCAACTGT 1035
Db      536 AGGCATCTAGGAAATGTGAGCACCCTACAGAAAGCCTGCATGCTGTGGCCAACTGT 477
Qy      1036 TTCACTGGAGCAAGAAAGAGATCTCATAGGACGGAGGGGAAATGTTTCCCTTCCAAGCT 1095
Db      476 TTCACTGGAGCAAGAAAGAGATCTCATAGGACGGAGGGGAAATGTTTCCCTTCCAAGCT 417
Qy      1096 TGGTCACTGTGTTAACTGCTTATCAGCTATTACACATCTCCATGCTTCTCCATGAAA 1155
Db      416 TGGTCACTGTGTTAACTGCTTATCAGCTATTACACATCTCCATGCTTCTCCATGAAA 357
Qy      1156 CTCTGTGTTTTCATCATCTCTTCTAGTTGACCTGCACAGCTTGTGTAGACCTAGATTTA 1215
Db      356 CTCTGTGTTTTCATCATCTCTTCTAGTTGACCTGCACAGCTTGTGTAGACCTAGATTTA 297
Qy      1216 ACCCTAAGTAAAGTGTCTGGGTATAGAACCGCTAAGAAATTTTCCCCCAAGGACTCTTGCT 1275
Db      296 ACCCTAAGTAAAGTGTCTGGGTATAGAACCGCTAAGAAATTTTCCCCCAAGGACTCTTGCT 237
Qy      1276 TCCTTAAGCCCTTCTGGCTTCTGTTTATGCTTCA-TTAAAGTATAAGCCCTTAACCTTGT 1334
Db      236 TCCTTAAGCCCTTCTGGCTTCTGTTTATGCTTCA-TTAAAGTATAAGCCCTTAACCTTGT 177
Qy      1335 CGCTAGTCTTAAGGAGAAACCTTTAAACCAAAAGTTTTTATCATTTGAACAAATTTGAA 1394
Db      176 CGCTAGTCTTAAGGAGAAACCTTTAAACCAAAAGTTTTTATCATTTGAACAAATTTGAA 117
Qy      1395 CAACCCCTATTTTGTGGGATGAGAA--GGGTGAATAGAGGCTTGGAGCTTTCCCTTT 1452
Db      116 CAACCCCTATTTTGTGGGATGAGAAACCGGGTGAATAGAGGCTTGGAGCTTTCCCTTT 57
Qy      1453 GTGTGGTAGGACTTGGAGGAGAAATCCCTGGACTTTTCACTAACCCCTGTGACATCTC 1510
Db      56 GTGTGGTAGGACTTGGAGGAGAA--TCCCTGGAGCTTCAATAAACCCCTGTGACATCTC 1

```

RESULT 11

```

B0690091
LOCUS      838 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_8064191 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6207831
           5', mRNA sequence.
ACCESSION B0690091
VERSION   B0690091.1 GI:21815407
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 838)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLC2364 row: i column: 16
High quality sequence stop: 670.

FEATURES
source

1..838
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6207831"
/clone_lib="NIH_MGC_110"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC library."

BASE COUNT 216 a 205 c 196 g 220 t 1 others
ORIGIN

Query Match 49.2%; Score 765.8; DB 14; Length 838;
Best Local Similarity 99.5%; Pred. No. 2.4e-228;
Matches 767; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 531 TTGAGCGCTATCTACTGATGTAGTACGCGGTACAAAGTGAGCAGCATCACCCCTCACCAGC 590
Db 1 TTGAGCGCTATCTACTGATGTAGTACGCGGTACAAAGTGAGCAGCATCACCCCTCACCAGC 60
Qy 591 AACTCCTTACCCTGATCTGTTCCAAAGTGGCAAGGAGGCAATCGCGGCCACAGATTG 650
Db 61 AACTCCTTACCCTGATCTGTTCCAAAGTGGCAAGGAGGCAATCGCGGCCACAGATTG 120
Qy 651 ACAAGAAAGGAGCGGCTGCTCATGACCTTCTCTGAGGAGATGTATCCGAGAATTTA 710
Db 121 ACAAGAAAGGAGCGGCTGCTCATGACCTTCTCTGAGGAGATGTATCCGAGAATTTA 180
Qy 711 ACTTAATGAGCTATACCAAGCGGGCCAGAAACTATCAAGGCTGGAGACAATATCCCTG 770
Db 181 ACTTAATGAGCTATACCAAGCGGGCCAGAAACTATCAAGGCTGGAGACAATATCCCTG 240
Qy 771 AGGAGCAGCCTGTGCTTCAACCCCAACCCAGCAGTGTCAAGTGGGAAACAGAGATA 830
Db 241 AGGAGCAGCCTGTGCTTCAACCCCAACCCAGCAGTGTCAAGTGGGAAACAGAGATA 300
Qy 831 AATAAGATCCTCACCTTTGGCAGTGTCTCTCTCTCTCAATTTCCAGGCTTTTCCATAAC 890
Db 301 AATAAGATCCTCACCTTTGGCAGTGTCTCTCTCTCTCAATTTCCAGGCTTTTCCATAAC 360
Qy 891 CACAAGCCTGAGGCTGCAGCCTTTTATTTATATTTTCCCTTTGGCTGTGACTGGGTGGG 950
Db 361 CACAAGCCTGAGGCTGCAGCCTTTTATTTATATTTTCCCTTTGGCTGTGACTGGGTGGG 420
Qy 951 CAGCATGAGCTTCTGATTTTAAAGAGGCAATCTAGGGAATTGTACAGCACCTTACAGGAA 1010
Db 421 CAGCATGAGCTTCTGATTTTAAAGAGGCAATCTAGGGAATTGTACAGCACCTTACAGGAA 480
Qy 1011 GGCCTGCCATGCTGTGGCCCACTGTTTCACTGGAGCAAGAAAGATCTCATAGGACGA 1070

```
Db 481 GGCCTGCCATGCTGTGGCCCACTGTTTCACTGGAGCAAGAAAGAGATCTCATAGGCGGA 540
QY 1071 GGGGGAATGGTTTCCCTCCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAG 1130
Db 541 GGGGGAATGGTTTCCCTCCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAG 600
QY 1131 ACATCTCCATGGTTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTCTAGTTGACCTG 1190
Db 601 ACATCTCCATGGTTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTCTAGTTGACCTG 660
QY 1191 CACAGCTTGGTTAGACTAGATTAAACCTTAAGCTAAGCTGCTGGGTATAGAACGCTAA 1250
Db 661 CACAGCTTGGTTAGACTAGATTAAACCTTAAGCTAAGCTGCTGGGTATAGAACGCTAA 720
QY 1251 GAATTTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCGCTTCGTTT 1300
Db 721 GAATTTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCGCTTCGTTT 770

RESULT 12
BQ962329
-^CUS
INITIATION BQ962329 835 bp mRNA linear EST 21-AUG-2002
AGENCOURT 8820581 NIH_MGC_42 Homo sapiens cdna clone IMAGE:6379790
5', mRNA sequence.
ACCESSION BQ962329
VERSION BQ962329.1 GI:22377807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 835)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cdna Library Preparation: Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LfCM2567 row: f column: 15
High quality sequence stop: 712.
Location/Qualifiers
1. .835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6379790"
/clone_lib="NIH_MGC_42"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cdna made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. !"
BASE COUNT 228 a 192 c 210 g 205 t
ORIGIN
Query Match 49.1%; Score 763.2; DB 14; Length 835;
Best Local Similarity 97.7%; Pred. No. 1.5e-227;
Matches 774; Conservative 18; Mismatches 22; Indels 0; Gaps 0;
QY 265 GATGTTTCTCAGTCCATGTGATGATGAGAACCGCAGATCCATCTCTCTGATGACGTG 324
|||||
```

```
Db 1 GATGTTTCTCAGTCCATGTGATGATGAGAACCGCAGATCCATCTCTCTGATGACGTG 60
QY 325 CAAACCCCCCTATATATATGGCCCTCAGTATATCAAGTACTTCAATGATAAAACCATTTGA 384
Db 61 CAAACCCCCCTATATATATGGCCCTCAGTATATCAAGTACTTCAATGATAAAACCATTTGA 120
QY 385 TGAGGAACCTAGAACGGGACAAAGAGGTCACCTTGGATTGTGGAGTTCTTTCCCAATTTGGTC 444
Db 121 TGAGGAACCTAGAACGGGACAAAGAGGTCACCTTGGATTGTGGAGTTCTTTCCCAATTTGGTC 180
QY 445 TAATGACTGCGCAATCATATTTGCCCTCTATCTATGCTGACCTCTCCCTTAAATACAACTGTAC 504
Db 181 TAATGACTGCGCAATCATATTTGCCCTCTATCTATGCTGACCTCTCCCTTAAATACAACTGTAC 240
QY 505 AGGCTAAATTTTGGGAAGTGGATGTTGGACGCTATCTATGATGTTAGTAGCGCGGTACAA 564
Db 241 AGGCTAAATTTTGGGAAGTGGATGTTGGACGCTATCTATGATGTTAGTAGCGCGGTACAA 300
QY 565 AGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCTGATCTCTGTTCCAAAGTGGCAA 624
Db 301 AGTGAGCACATCACCCCTCACCAAGCAACTCCCTACCTGATCTCTGTTCCAAAGTGGCAA 360
QY 625 GGAGGCAATCGCGGGCCACAGATTGACAAAGAGGAGCGGCTGTCTCATGGACCTTCTC 684
Db 361 GGAGGCAATCGCGGGCCACAGATTGACAAAGAGGAGCGGCTGTCTCATGGACCTTCTC 420
QY 685 TGAGGAGAATGTGATCCGGAATTTAACTTAAATGAGCTATACCAGCGGGGCCAAGAACT 744
Db 421 TGAGGAGAATGTGATCCGGAATTTAACTTAAATGAGCTATACCAGCGGGGCCAAGAACT 480
QY 745 ATCAAAGGCTGGAGACAAATATCCCTGAGGAGAGCCCTGTGGCTTCAACCCACACACAGT 804
Db 481 ATCAAAGGCTGGAGACAAATATCCCTGAGGAGAGCCCTGTGGCTTCAACCCACACAGT 540
QY 805 GTCAGATGGGGAACAAAGAGGATAAATAAGATTCCTCAGTTTGGCAGTGTCTCTCTCC 864
Db 541 GTCAGATGGGGAACAAAGAGGATAAATAAGATTCCTCAGTTTGGCAGTGTCTCTCTCC 600
QY 865 TGTCATTTCCAGGCTCTTCCATTAACCAACCAAGCTGAGGCTGAGGCTTTTATTTATGTT 924
Db 601 TGTCATTTCCAGGCTCTTCCATTAACCAACCAAGCTGAGGCTGAGGCTTTTATTTATGTT 660
QY 925 TTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGTCAGCTTCTGATTTTAAAGAGCATCTA 984
Db 661 TTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGTCAGCTTCTGATTTTAAAGAGCATCTA 720
QY 985 GGAATTTGTCAGGACCCCTACAGGAAGCCCTGCCATCTGTGGCCAACTGTTTCACTGGA 1044
Db 721 GGAATTTGTCAGGACCCCTACAGGAAGCCCTGCCATCTGTGGGCCAACTGTTTCACTG 780
QY 1045 GCAAGAAAGAGA 1056
Db 781 GGAGCAAGGAAA 792

RESULT 13
BQ920164
LOCUS BQ920164 913 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8918912 NIH_MGC_40 Homo sapiens cdna clone IMAGE:6383675
5', mRNA sequence.
ACCESSION BQ920164
VERSION BQ920164.1 GI:22334862
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 913)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: DCTD/DTP
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2577 row: h column: 12
 High quality sequence stop: 684.
 Location/Qualifiers
 1. .913
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6383675"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dt priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 234 a 241 c 215 g 221 t 2 others
 ORIGIN

Query Match 48.3%; Score 751.6; DB 14; Length 913;
 Best Local Similarity 92.7%; Pred. No. 6.9e-224;
 Matches 844; Conservative 0; Mismatches 31; Indels 35; Gaps 4;

QY 51 CCGAAAGATGGCGGCTTGGACCTCTAATTCCTGCTGTTATTCGGTCCGGGACTTT 110
 DB 1 CCGAAAGATGGCGGCTTGGACCTCTAATTCCTGCTGTTATTCGGTCCGGGACTTT 60
 QY 111 CAGATGGCTGGCCCAACCTTACTACCTTCTGCGGCCCTGCTCTGCTGCTTCCTAC 170
 DB 61 CAGATGGCTGGCCCAACCTTACTACCTTCTGCGGCCCTGCTCTGCTGCTTCCTAC 120
 QY 171 TCGTGAGGAACCTGGCGGCTCTGCCACGGTCTGCCACCCACCGGAGAGCGT---- 226
 DB 121 TCGTGAGGAACCTGGCGGCTCTGCCACGGTCTGCCACCCACCGGAGAGCGTCTTA 180
 C 227 -----AACCGTGTGACTTTGACTGGAGAGAGTGGAGA 260
 181 CAGATGCACCCCATGCTGGAGACGCAACCCCTTTGACAGAAAGAGAGAGTGGAGA 240
 261 TCCTGATGTTTCTCAGTGCCATTGTGATGATGAAGAACCCGACATCCATGTTCTGATGA 320
 DB 241 TCCTGATGTTTCTCAGTGCCATTGTGATGATGAAGAACCCGACATCCATGTTCTGATGA 300
 QY 321 CGTGCAACCCCCCTATATATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAACCA 380
 DB 301 CGTGCAACCCCCCTATATATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAACCA 360
 QY 381 TTGATGAGGAACCTAGAACGGGCAAGAGGTCTACTTGGATTGCGAGTCTTTGCCAATT 440
 DB 361 TTGATGAGGAACCTAGAACGGGCAAGAGGTCTACTTGGATTGCGAGTCTTTGCCAATT 420
 QY 441 GGTCTAATGACTGCCAATCATTTGCCCTTATCTATGCTGACCTCTCCCTTAAATACAAC 500
 DB 421 GGTCTAATGACTGCCAATCATTTGCCCTTATCTATGCTGACCTCTCCCTTAAATACAAC 480
 QY 501 GTACAGGCGTAATTTTGGGAGGTGGATGTTGGAGCGCTATACCTGATGTAGTACGCGGT 560
 DB 481 GTACAGGCGTAATTTTGGGAGGTGGATGTTGGAGCGCTATACCTGATGTAGTACGCGGT 540
 QY 561 ACAAGTGGACACATCACCCCTCACCAGCAACTCCCTACCTGATGCTTCCCAAGGTG 620
 DB 541 ACAAGTGGACACATCACCCCTCACCAGCAACTCCCTACCTGATGCTTCCCAAGGTG 600
 QY 621 GCAAGGAGCAATGCGGGGCCACAGATTGACAAGAAAGACGGGCTGTCTCATGACCT 680

DB 601 GCAAGGAGCAATGCGGGGCCACAGATTGACAAGAGGCGGCTGTCTCATGACCT 660
 QY 681 TCTCTGAGGAGAAATGATCCGAGAAATTTAACTAAATGAGCTATACACGCGGCAAGA 740
 DB 661 TCTCTGAGGAGAAATGATCCGAGAAATTTAACTAAATGAGCTATACACGCGGCAAGA 720
 QY 741 AACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACA 800
 DB 721 AACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACA 780
 QY 801 CAGTGTCAAGTGGGAAAACAAGGATTAATTAAGATCCCTCACTTTGGCAGTCTCTCC 859
 DB 781 CAGTGTCAAGTGGGAAAACAAGGATTAATTAAGATCCCTCACTTTGGCAGGCTTC 840
 QY 860 TCT---CCTGTCAATTCAGGCTCTTCCATAACCAACA--GCCTGAGGCTGCAGCCTTT 915
 DB 841 CTTCTCCCTGCTATTCAGGCTCTTCCATAACCAACAAGGCTGAGGCTGCAGNCCCT 900
 QY 916 ATTTATGTTT 925
 DB 901 TTTTATTAT 910

RESULT 14
 LOCUS BQ687916 894 bp mRNA linear EST 15-JUL-2002
 DEFINITION AGENCOURT_8344073 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6248661
 5', mRNA sequence.
 ACCESSION BQ687916
 VERSION BQ687916.1 GI:21813232
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 894)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabps-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2388 row: n column: 22
 High quality sequence stop: 645.
 Location/Qualifiers
 1. .894
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6248661"
 /clone_lib="NIH_MGC_110"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dt priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 236 a 204 c 222 g 231 t 1 others
 ORIGIN

FEATURES
 source

Query Match 48.3%; Score 750.8; DB 14; Length 894;
 Best Local Similarity 99.1%; Pred. No. 1.2e-223;
 Matches 807; Conservative 0; Mismatches 2; Indels 5; Gaps 5;

QY	369	ATGATAAAACCATGATGAGGAACCTAGAACGGGACAAAGAGGCTCACTTGGATTGTGGAGT	428
Db	1	ATGATAAAACCATGATGAGGAACCTAGAACGGGACAAAGAGGCTCACTTGGATTGTGGAGT	60
QY	429	TCTTTGCCAATTGGTCTAATGACTGCCAATCAATTTGGCCCTATCTATGCTGACCTCTCC	488
Db	61	TCTTTGCCAATTGGTCTAATGACTGCCAATCAATTTGGCCCTATCTATGCTGACCTCTCC	120
QY	489	TTAATAACAACGTACAGGGCTAAATTTTGGGAAGTGGGATCTTGGACGCTATACCTGATG	548
Db	121	TTAATAACAACGTACAGGGCTAAATTTTGGGAAGTGGGATCTTGGACGCTATACCTGATG	180
QY	549	TTAGTACCGGTACAAAAGTGGACACATCACCCCTCACCAAGCACTCCCTACCTGATCC	508
Db	181	TTAGTACCGGTACAAAAGTGGACACATCACCCCTCACCAAGCACTCCCTACCTGATCC	240
QY	609	TGTTCCAAAGTGGCAAGGAGCAATGCGGCGCCACAGATTACCAAGAAAGACGGGCTG	668
Db	241	TGTTCCAAAGTGGCAAGGAGCAATGCGGCGCCACAGATTACCAAGAAAGACGGGCTG	300
QY	669	TCTCATGACCTTCTCTGAGGAGATGTGATCCGAGAAATTTAACTTAAATGAGCTATACC	728
Db	301	TCTCATGACCTTCTCTGAGGAGATGTGATCCGAGAAATTTAACTTAAATGAGCTATACC	360
QY	729	AGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTT	788
Db	361	AGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTT	420
QY	789	CAACCCCCACACACTGTCAGATGGGGAACAAGAGGATAAATAAGATCCTCACTTTG	848
Db	421	CAACCCCCACACACTGTCAGATGGGGAACAAGAGGATAAATAAGATCCTCACTTTG	480
QY	849	GCAGTGTCTCTCTCTGTCATATCCAGGCTCTTCCATACCAAGCCTGAGGCTGCA	908
Db	481	GCAGTGTCTCTCTCTGTCATATCCAGGCTCTTCCATACCAAGCCTGAGGCTGCA	540
QY	909	GCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGTGGGCGAGCATCGCTTCTGAT	968
Db	541	GCCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGTGGGCGAGCATCGCTTCTGAT	600
QY	969	TTTAAAGAGGATCTAGGGAATTTGTCAGGACCCCTACAGGAAGGCTGCATGCTGTGG	1028
Db	601	TTTAAAGAGGATCTAGGGAATTTGTCAGGACCCCTACAGGAAGGCTGCATGCTGTGG	660
QY	1029	CAACTGTTTCTACTGGAGCAAGAGATCTCATAGGACGGA-GGGGGAAATGGTTTCCC	1087
Db	661	CAACTGTTTCTACTGGAGCAAGAGATCTCATAGGACGGAAGGAAATGGTTTCCC	720
QY	1088	TCCAAAGCTT-GGGTCAGTGTGTTAACT-GCTTATCAGCTATTACAGACATCTCCATGGT	1145
Db	721	TCCAGGCTTGGGTCAGTGTGTTAACTGCTTATCAGCTATTACAAACATCTCCATGGT	780
QY	1146	CTCCATGAA-ACTCTGT-GGTTTCATCTTCCTT	1177
Db	781	CTCCATGAANACTCTGTGGGTTTCATCACTTCCTT	814
RESULT	15		
LOCUS	AL569941/c		
DEFINITION	AL569941 LTI_NFL006_PL2 Homo sapiens cdna clone CS0DI003YH21 3		
ACCESSION	AL569941		
VERSION	AL569941.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.		
	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES	Location/Qualifiers 1..786 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0DI003YH21" /clone_lib="LTI_NFL006_PL2" /tissue_type="placenta" /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	226 a 181 c 171 g 204 t		
ORIGIN			
Query Match	48.0%; Score 746.2; DB 9; Length 786;		
Best Local Similarity	98.9%; Pred. No. 3.2e-222;		
Matches	778; Conservative 4; Mismatches 2; Indels 3; Gaps 3;		
QY	677	ACCTCTCTCTGAGGAGAAATGTGATCCGAGAAATTTAACTTAAATGAGCTATACCGCGGCC	736
Db	786	MCCTTTCTGAGGAGAAATGTGATCCGAGAAATTTAACTTAAATGAGCTATACCGCGGCC	727
QY	737	AAGAAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACCC	796
Db	726	AAGAAACTATMAAGGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACCC	667
QY	797	ACCACAGTCTCAGATGGGAAAACAAGAGGATAAATAAGATCCTCACATTTGGCAGTCT	856
Db	666	ACCACAGTCTCAGATGGGAAAACAAGAGGATAAATAAGATCCTCACATTTGGCAGTCT	607
QY	857	TCTCTCTCTCTCAATTTCCAGGCTCTTTCCATACCAAGCCTGAGGCTCAGCCTTTTA	916
Db	606	TCTCTCTCTCTCAATTTCCAGGCTCTTTCCATACCAAGCCTGAGGCTCAGCCTTTTA	547
QY	917	TTATGTTTTTCCCTTTGGCTGTGACTGGTGGGCGAGCATGACGCTTCTGATTTTAAAGA	976
Db	546	TTATGTTTTTCCCTTTGGCTGTGACTGGTGGGCGAGCATGACGCTTCTGATTTTAAAGA	487
QY	977	GGCATCTAGGGAATGTCAGGCACCTCAGGAAGGCTGCCATGCTGTGGCCAACTGTT	1036
Db	486	GGCATCTAGGGAATGTCAGGCACCTCAGGAAGGCTGCCATGCTGTGGCCAACTGTT	427
QY	1037	TCACCTGAGCAAGAAAGAGATCTATAGGAGGAGGGGAAATGGTTTCCCTCCAAGCTT	1096
Db	426	TCACCTGAGCAAGAAAGAGATCTATAGGAGG- GGGGAAATGGTTTCCCTCCAAGCTT	368
QY	1097	GGGTAGTGTGTTAACTGCTTATCAGCATCTTCCATGCTTCTCCATGAAC	1156
Db	367	GGGTAGTGTGTTAACTGCTTATCAGCATCTTCCATGCTTCTCCATGAAC	308
QY	1157	TCGTGGTTTCATCATCTCTTCTAGTTCACCTGACGAGCTTGGTTAGACCTAGATTTAA	1216
Db	307	TCGTGGTTTCATCATCTCTTCTAGTTCACCTGACGAGCTTGGTTAGACCTAGATTTAA	248
QY	1217	CCCTAAGGTAAGATGCTGGGGTATACACGCTAAGAAATTTCCCCCAAGGACTCTTGCCT	1276
Db	247	CCCTAAGGTAAGATGCTGGGGTATACACGCTAAGAAATTTCCCCCAAGGACTCTTGCCT	188
QY	1277	CCTTAAGCCCTCTCTGGCTTTCGT-TTATGTTCTTATTAAAGATTAAGCCTTAACCTTGC	1335
Db	187	CCTTAAGCCCTCTCTGGCTTTCGT-TTATGTTCTTATTAAAGATTAAGCCTTAACCTTGC	128

Qy	1336	GCTAGTCCTAAGGAGAAACCTTTAACCAAGTTTATCATTTGAAGACAATATTGAAC	1395
Db	127	GCTAGTCCTAAGGAGAAACCTTTAACCAAGTTTATCATTTGAAGACAATATTGAAC	68
Qy	1396	AACCCCTAT-TTTGTGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCCTTTGT	1454
Db	67	AACCCCTATATTGTGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCCTTTGT	8
Qy	1455	GTGCTAG 1461	
Db		7 GTGCTAG 1	

Search completed: July 12, 2003, 11:06:13
 Job time : 1612 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 12, 2003, 08:35:42 ; Search time 2787 Seconds
(without alignments)
16237.841 Million cell updates/sec

Title: US-09-954-846-4
Perfect score: 1555
Sequence: 1 AGGGGAGGGGGCGGAGACC.....TAATAAAGATTGGGATTA 1555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1381.2	88.8	1644	9	AF059753	AF059753 Homo sapi
2	1379.4	88.7	1646	9	BC000666	BC000666 Homo sapi
3	1370.2	88.1	1631	6	AX136115	AX136115 Sequence
4	1370	88.1	1620	6	AX375964	AX375964 Sequence
5	1355.4	87.2	1669	9	AF132965	AF132965 Homo sapi
6	962.2	61.9	189271	9	AL138752	AL138752 Human DNA
7	934.2	60.1	112389	9	AC073598	AC073598 Homo sapi
8	868	55.8	110116	9	AP001931	AP001931 Homo sapi
9	868	55.8	130642	2	AP001082	AP001082 Homo sapi
10	868	55.8	158349	2	AC009438	AC009438 Homo sapi
11	868	55.8	161397	2	AC090218	AC090218 Homo sapi
12	868	55.8	170540	2	AC092707	AC092707 Homo sapi
13	868	55.8	199992	2	AC021522	AC021522 Homo sapi
14	856	55.0	153394	2	AP000727	AP000727 Homo sapi
15	571.6	36.8	166341	2	AC013381	AC013381 Homo sapi
16	548.4	35.3	1978	10	BC013544	BC013544 Mus muscu
17	544.2	35.0	200653	2	AC127695	AC127695 Mus muscu
18	463	29.8	726	6	AX136467	AX136467 Sequence
19	328.8	21.1	395	11	GI4613	GI4613 human STS S
20	258.4	16.6	307	6	AX396397	AX396397 Sequence
21	252.2	16.2	5423	6	AX336311	AX336311 Sequence
22	252.2	16.2	5423	9	AB002382	AB002382 Human mRN
23	246	15.8	117711	9	AP000662	AP000662 Homo sapi
24	246	15.8	161397	2	AC090218	AC090218 Homo sapi
25	221.4	14.2	1730	9	BC008500	BC008500 Homo sapi
26	158.6	10.2	183656	2	AC111385	AC111385 Rattus no
27	156.2	10.0	149481	2	AC102678	AC102678 Mus muscu
28	156.2	10.0	216237	2	AC121786	AC121786 Mus muscu
29	153.4	9.9	81994	2	AC096003	AC096003 Rattus no
30	143.2	9.2	183656	2	AC111385	AC111385 Rattus no
31	127	8.2	127	11	G62115	G62115 1897 Human
32	110.2	7.1	153394	2	AP000727	AP000727 Homo sapi
33	95.8	6.2	1223	3	AY118488	AY118488 Drosophil
34	95.8	6.2	31857	2	AC020299	AC020299 Drosophil
35	95.8	6.2	45105	3	AC004311	AC004311 Drosophil
36	95.8	6.2	160106	3	AC099014	AC099014 Drosophil
37	95.8	6.2	173347	3	AC099013	AC099013 Drosophil
38	95.8	6.2	248096	3	AE003795	AE003795 Drosophil
39	89.6	5.8	81994	2	AC096003	AC096003 Rattus no
40	53.6	3.4	36882	3	U21324	U21324 Caenorhabdi
41	43.2	2.8	161888	9	HS70501	AL035251 Human DNA
42	42.6	2.7	14461	6	AX256438	AX256438 Sequence
43	42.4	2.7	5204	5	GGU58204	U58204 Gallus gall
44	41.8	2.7	67446	2	AC103818	AC103818 Homo sapi
45	41.8	2.7	184302	9	AL162393	AL162393 Human DNA

ALIGNMENTS

RESULT 1
AF059753
LOCUS AF059753 1644 bp mRNA linear PRI 02-JAN-2001
DEFINITION Homo sapiens clone 008a05 My009 protein mRNA, complete cds.
ACCESSION AF059753
VERSION AF059753.1 GI:12001957
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1644)
AUTHORS Mao,Y.M., Xie,Y., Zhou,Z.X., Ying,K. and Zheng,Z.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Institute of Genetics, School of Life

Science, Fudan University, 220 Handan Rd., Shanghai 200433,
P.R.China

FEATURES Location/Qualifiers
source 1..1644

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="008a05"
/tissue_type="brain"
/dev_stage="fetus"
28..1146
/codon_start=1
/product="My009 protein"
/protein_id="A04312.1"
/db_xref="GI:12001958"

/translation="MAVLAPLIALVYVPRLSRWLAQPYLLSALLSRAFLVLRKLPP
LCHGLDREDFDPCDFDREVEILMELSAIVMMKNRRSITVEQHGIFNMFQSVANT
ILFLRLDRLMLGLITLGLVFLMTKPLPYLMPEYIKYFNDRKTIIDEEELKRVWIV
EFFANWSDCOSFAPYADLSKYNCTGLNFKGVDRVTVSTYKSTSLTKOLP
TLILPOGKEARRRPOISKKGRAVSWTFSEENVIREFNELNELYORAKLSKAGDNIPE
EQPVASTPPQCMGKTRINKILITLAVLPLLSIPGSFHNKPEAAAFYLCFFPGCDWV
GQHAASDFKASRELSGLQELPCCGLFHWKREIS"

BASE COUNT 412 a 401 c 385 g 446 t

Juery Match 88.8%; Score 1381.2; DB 9; Length 1644;

Best Local Similarity 92.8%; Pred. No. 0; Mismatches 3; Indels 115; Gaps 2;
Matches 1518; Conservative 0;

Qy 33 GAGCAGTGGCGGCTTACGGCGAAAGATGGCGGTCTTGGCACCTCTAAATTGCTCGTGT 92
Db 2 GAGCAGTGGCGGCTTACGGCGAAAGATGGCGGTCTTGGCACCTCTAAATTGCTCGTGT 61
Qy 93 ATTCGGTGGCGGACTTTCACGATGGCTCGGCCAACCTTACTACCTTCTGTGGCCCTGC 152
Db 62 ATTCGGTGGCGGACTTTCACGATGGCTCGGCCAACCTTACTACCTTCTGTGGCCCTGC 121
Qy 153 TCTCTGCTGCTTCTACTCGTGGAGAACTCGCGCGCTCTGCCACGCTCTGCCACCC 212
Db 122 TCTCTGCTGCTTCTACTCGTGGAGAACTCGCGCGCTCTGCCACGCTCTGCCACCC 181
Qy 213 AACGGAAGACGGTAAACCGCTGTACTTTGACTGGAGAGAAGTGGAGATCTCTGATTTTC 272
Db 182 AACGGAAGACGGTAAACCGCTGTACTTTGACTGGAGAGAAGTGGAGATCTCTGATTTTC 241
Qy 273 TCAGTGCCATTGTGATGATGAAGAACCGCAGATCCA ----- 308
Db 242 TCAGTGCCATTGTGATGATGAAGAACCGCAGATCCA ----- 301
Qy 309 ----- 308
Db 302 ACATTTTCATGTTAGTAAAGTGGCCACACAAATTTCTTTCTCCCTTGGATATTCGCA 361
Qy 309 -----TCTCTGATGACGTGCAACCCCTCTAT 338
Db 362 TGGGCTACTTTACATCACACTCTGCATAGTGTCTCTGATGACGTGCAACCCCTCTAT 421
Qy 339 ATATGGGCGCTGAGTATATCAAGTACTTCAATGATAAAACCATTGATGAGGAACAGAAC 398
Db 422 ATATGGGCGCTGAGTATATCAAGTACTTCAATGATAAAACCATTGATGAGGAACAGAAC 481
Qy 399 GGGACAAGAGGTCACCTGAGTGTGGAGTCTTTCGCAATTTGGCTAATGACTGCCAAT 458
Db 482 GGGACAAGAGGTCACCTGAGTGTGGAGTCTTTCGCAATTTGGCTAATGACTGCCAAT 541
Qy 459 CATTTGGCCCTATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTG 518
Db 542 CATTTGGCCCTATCTATGCTGACCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTTG 601
Qy 519 GGAAGTGGATGTTGGACGCTTACTGATGTTAGTACGGGTCACAAAGTGAAGACATCAC 578
Db 602 GGAAGTGGATGTTGGACGCTTACTGATGTTAGTACGGGTCACAAAGTGAAGACATCAC 661
Qy 579 CCCTACCAAGCAACTCCCTACCTGATCTCTTCCAGGTGGCAAGGAGCAATCGGC 638

RESULT 2
BC000666
LOCUS
DEFINITION

BC000666 1646 bp mRNA linear PRI 12-JUL-2001
Homo sapiens, CGI-31 protein, clone MGC:713 IMAGE:3349733, mRNA,
complete cds.

Db 662 CCCTCACCAGCAACTCCCTACCTGATCTGTTCACAGGTGCAGAGGCAATCGGC 721
Qy 639 GGGCAGAGATTGACAAAGAGGACGGCTGTCTCATGGACCTTCTCTGAGGAGAATGTGA 698
Db 722 GGGCAGAGATTGACAAAGAGGACGGCTGTCTCATGGACCTTCTCTGAGGAGAATGTGA 781
Qy 699 TCCGAGAAATTTAAATGAGCTATACGAGGGGCCAAGAACTATCAAGAGCTGGAG 758
Db 782 TCCGAGAAATTTAAATGAGCTATACGAGGGGCCAAGAACTATCAAGAGCTGGAG 841
Qy 759 ACAATATCCTGAGGAGCAGCTGTGGCTTCAACCCCCACACAGTGTCTAGATGGGAAA 818
Db 842 ACAATATCCTGAGGAGCAGCTGTGGCTTCAA -CCCCACCAGTGTCTAGATGGGAAA 900
Qy 819 ACAAGAGGATTAATAAGATCTCTACTTTGGGAGTCTCTCTCTCTCTCAATTCAGGC 878
Db 901 ACAAGAGGATTAATAAGATCTCTACTTTGGGAGTCTCTCTCTCTCTCAATTCAGGC 960
Qy 879 TCTTTCCATTAACCAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 938
Db 961 TCTTTCCATTAACCAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 1020
Qy 939 GACTGGTGGGCGCAGCATGCAGCTTCTGATTTTAAAGAGCATCTAGGGAATGTCAAGC 998
Db 1021 GACTGGTGGGCGCAGCATGCAGCTTCTGATTTTAAAGAGCATCTAGGGAATGTCAAGC 1080
Qy 999 ACCCTACAGAAAGGCTGCGCATGCTGTGCCAACTGTTTCACTGGAGCAAGAAAGATC 1058
Db 1081 ACCCTACAGAAAGGCTGCGCATGCTGTGCCAACTGTTTCACTGGAGCAAGAAAGATC 1140
Qy 1059 TCATAGAGCGGAGGGGAAATGGTTTCCCTCCAAAGCTTGGGTGAGTGTCTTAAGTCTTA 1118
Db 1141 TCATAGAGCGGAGGGGAAATGGTTTCCCTCCAAAGCTTGGGTGAGTGTCTTAAGTCTTA 1200
Qy 1119 TCAGCTATTTCAGACATCTCCATGTTTCCCATGAACTCTGTGGTTTCAATCTCTTC 1178
Db 1201 TCAGCTATTTCAGACATCTCCATGTTTCCCATGAACTCTGTGGTTTCAATCTCTTC 1260
Qy 1179 TTAGTTGACCTGCACAGCTTGGTTAGACCTAGATTTTAAACCTTAAGGTAAGATGCTGGGT 1238
Db 1261 TTAGTTGACCTGCACAGCTTGGTTAGACCTAGATTTTAAACCTTAAGGTAAGATGCTGGGT 1320
Qy 1239 ATAGAACGCTAAGAAATTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCTGCTTCGT 1298
Db 1321 ATAGAACGCTAAGAAATTTCCCAAGGACTCTTGTCTTCCCTTAAGCCCTTCTGCTTCGT 1380
Qy 1299 TTATGCTCTTCAATTAAGATTAAGCCTTAACCTTGTGCTAGTCTTAAGGAGAAACCTTT 1358
Db 1381 TTATGCTCTTCAATTAAGATTAAGCCTTAACCTTGTGCTAGTCTTAAGGAGAAACCTTT 1440
Qy 1359 AACCAAAAGTTTATCATTTGAAGACAATATTGAACACCCCTATTTTGTGGGATG 1418
Db 1441 AACCAAAAGTTTATCATTTGAAGACAATATTGAACACCCCTATTTTGTGGGATG 1500
Qy 1419 AGAAGGGTGAATAGAGCTTGAAGCTTTCCTTTTGTGTGTGAGGACTTGGAGAGAATC 1478
Db 1501 AGAAGGGTGAATAGAGCTTGAAGCTTTCCTTTTGTGTGTGAGGACTTGGAGAGAATC 1560
Qy 1479 CCTGACCTTTCACCTAACCTCTGACATCTCCCAACACCCAGTGTGATGGCTTCCGTAA 1538
Db 1561 CCTGACCTTTCACCTAACCTCTGACATCTCCCAACACCCAGTGTGATGGCTTCCGTAA 1620
Qy 1539 TAAAAAGATTGGGATT 1554
Db 1621 TAAAAAGATTGGGATT 1636

ACCESSION	BC000666	QY	170	CTCGTGAGAAACGCGCCGCTCTGCGACAGTCTGCGCCACCCACCGAAGACGGTAC	229
VERSION	BC000666.1	Db	121	CTCGTGAGAAACGCGCCGCTCTGCGACAGTCTGCGCCACCCACCGAAGACGGTAC	180
KEYWORDS	MGC.	QY	230	CCGTGTGACTTTGACTGGAGAGAAAGTGGAGATCCTGTGATGTTTCTCAGTGCCCATTTGTGATG	289
SOURCE	Homo sapiens.	Db	181	CCGTGTGACTTTGACTGGAGAGAAAGTGGAGATCCTGTGATGTTTCTCAGTGCCCATTTGTGATG	240
ORGANISM	Homo sapiens	QY	290	ATGAAGAACCAGCATGCA-----	308
REFERENCE	Submitted (15-NOV-2000)	Db	241	ATGAAGAACCAGCATGCAATTCATCTGCGGAGCAACATATAGGCAACATTTTCATGTTTAGT	300
AUTHORS	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	QY	309	-----	308
TITLE	NIH-MGC Project URL: http://mgs.nci.nih.gov	Db	301	AAAGTGGCCAAACAAATTTCTTCTCGCTTGGATATTCGCATGGGCTACTTTACATC	360
JOURNAL	Contact: MGC help desk	QY	309	-----TGTTCTCTGATGACGTGCAAAACCCCTCTATATATGGCCCTCAGTAT	355
	Email: gcapbs@mail.nih.gov	Db	361	ACACTCTGCATAGTGTCTCTGATGACGTGCAAAACCCCTCTATATATGGCCCTCAGTAT	420
	Tissue Procurement: ATCC	QY	356	ATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACATAGAACGGGCAAGAGGCTACT	415
	cDNA Library Preparation: Rubin Laboratory	Db	421	ATCAAGTACTTCAATGATAAAACCATTTGATGAGGAACATAGAACGGGCAAGAGGCTACT	480
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	QY	416	TGGATTCTGGAGTTCCTTGGCAATTTGGTCTTAATGACTCCCAATCATTTGCCCTCTATCTAT	475
	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;	Db	481	TGGATTCTGGAGTTCCTTGGCAATTTGGTCTTAATGACTCCCAATCATTTGCCCTCTATCTAT	540
	Web site: http://www.nisc.nih.gov/	QY	476	GCTCAGCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTGGGAAGTGGATGTTTGA	535
	Contact: nisc.mgc@nih.gov	Db	541	GCTCAGCTCTCCCTTAAATACAACTGTACAGGGCTAAATTTGGGAAGTGGATGTTTGA	600
	Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.	QY	536	CGCTATCTGATGTAGTACGCGGTACAAAGTGGAGCAATCACCCTCACCAGCAACTC	595
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	Db	601	CGCTATCTGATGTAGTACGCGGTACAAAGTGGAGCAATCACCCTCACCAGCAACTC	660
	Series: IRAL Plate: 5 Row: d Column: 22	QY	596	CCTACCTCTGATCTCTTCCAAAGTGGCAAGAGGCAATGCGGCGGCCACAGATTGACAAG	655
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7705725.	Db	661	CCTACCTCTGATCTCTTCCAAAGTGGCAAGAGGCAATGCGGCGGCCACAGATTGACAAG	720
FEATURES	Location/Qualifiers	QY	656	AAAGGACGGGTCTCTCATGACCTCTCTCTGAGAGAAATGATCCGAGAAATTTAACTTA	715
source	1. .1646	Db	721	AAAGGACGGGTCTCTCATGACCTCTCTCTGAGAGAAATGATCCGAGAAATTTAACTTA	780
	/organism="Homo sapiens"	QY	716	AATGAGCTATACACGCGGCGCAAGAAATATCAAGGCTGGAGACAATATCCCTGAGGAG	775
	/db_xref="LocusID:51075"	Db	781	AATGAGCTATACACGCGGCGCAAGAAATATCAAGGCTGGAGACAATATCCCTGAGGAG	840
	/db_xref="taxon:9606"	QY	776	CAGCCTGTGGCTTCAACCCCGCCACACAGTGTACAGTGGGAAACACAGAGGATAAATAA	835
	/clone="MGC:713 IMAGE:3349733"	Db	841	CAGCCTGTGGCTTCAACCCCGCCACACAGTGTACAGTGGGAAACACAGAGGATAAATAA	900
	/tissue.type="Eye, retinoblastoma"	QY	836	GATTCCTCAGTTTGGCAGTGTCTCTCTCTGATCAATTCAGGCTCTTCCATATACCAACAA	895
	/clone_lib="NIE_MGC_16"	Db	901	GATTCCTCAGTTTGGCAGTGTCTCTCTCTGATCAATTCAGGCTCTTCCATATACCAACAA	960
	/lab_host="DH10B-R"	QY	896	GCTGAGGCTGCGACCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGACGA	955
	/note="Vector: pOTB7"	Db	961	GCTGAGGCTGCGACCTTTTATTTATGTTTCCCTTTGGCTGTGACTGGGTGGGCGACGA	1020
	10. .900	QY	956	TGCAGCTTCTGATTTTAAAGAGGCAATCAGGAAATTCAGGCAACCTACAGGAAGGCCT	1015
	/product="CGI-31 protein"	Db	1021	TGCAGCTTCTGATTTTAAAGAGGCAATCAGGAAATTCAGGCAACCTACAGGAAGGCCT	1080
	/protein_id="AAH00666.1"	QY	1016	GCCATCTGTGGCCAACTGTTTTCACCTGGAGCAAGAGAGATCTCATAGGACGAGGGG	1075
	/db_xref="GI:12653761"	Db	1081	GCCATCTGTGGCCAACTGTTTTCACCTGGAGCAAGAGAGATCTCATAGGACGAGGGG	1140
	/translation="MAVLAPLALVSVPRLSRWLAOPYLLSALLSAAFLVLRKLPP LCHGLPTQREDNCPDFWREVEILFLSAIVMKNRISITVEQHTGNIFMFSKVANT ILFLPLDIRMGSLITLCLVFLMTKPLMGPEYIKYFNDKTIDELERDRKRTWIV EFFANWSDCQFAPYADLSLKYNCTGLNFKGVGVYDVTYKRVKVSPLTKOLP TLILFQSGKEANRRQIDKKGRVSWTFSEENVIREFNELNYQRAKLKSGADNPIE EQPVASTTTVSDGKGNKDK"	QY	1076	AAATGGTTTCCCTCCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATC	1135
	429 a 395 c 376 g 446 t	Db	1141	AAATGGTTTCCCTCCCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATC	1200
BASE COUNT	429 a 395 c 376 g 446 t				
ORIGIN					
Query Match	88.7%;				
Best Local Similarity	92.9%;				
Matches 1504;	Conservative 0; Mismatches 1; Indels 114; Gaps 1;				
QY	50. GCCGAAAAGATGGCGCTCTTGGCAGCTCTAAATGCTCTGCTGATTCGGTGGCGGACATT		109		
Db	1 GCCGAAAAGATGGCGCTCTTGGCAGCTCTAAATGCTCTGCTGATTCGGTGGCGGACATT		60		
QY	110 TCAGGATGGCTGGCCCAACCTTACTACCTTCTGTCGGCCCTGCTCTCTGCTGCTTCCCTA		169		
Db	61 TCAGGATGGCTGGCCCAACCTTACTACCTTCTGTCGGCCCTGCTCTCTGCTGCTTCCCTA		120		

```
QY 1136 TCCATGGTTCTCCATGAACCTCTGTGTTTCATCATCTCTTCTAGTTGACCTGCACAG 1195
Db 1201 TCCATGGTTCTCCATGAACCTCTGTGTTTCATCATCTCTTCTAGTTGACCTGCACAG 1260
QY 1196 CTGGTTAGACCTAGATTTAACCCCTAAGTAAGATGCTGGGTATAGAACGCTAAGAATT 1255
Db 1261 CTGGTTAGACCTAGATTTAACCCCTAAGTAAGATGCTGGGTATAGAACGCTAAGAATT 1320
QY 1256 TTCCCCAAGGACTCTTGGTTCCTTAAGCCCTTCTGGCTTCGCTTTATGTCCTTCAATAA 1315
Db 1321 TTCCCCAAGGACTCTTGGTTCCTTAAGCCCTTCTGGCTTCGCTTTATGTCCTTCAATAA 1380
QY 1316 AGTATAAGCCTAACCTTGTGCTAGCTCTAAGGAGAAACCTTTAACCACAAAGTTTAT 1375
Db 1381 AGTATAAGCCTAACCTTGTGCTAGCTCTAAGGAGAAACCTTTAACCACAAAGTTTAT 1440
QY 1376 CATTGAAGACAATATTGAACACCCCTATTTTGTGGGATTGAGAAGGGTGAATAGAG 1435
Db 1441 CATTGAAGACAATATTGAACACCCCTATTTTGTGGGATTGAGAAGGGTGAATAGAG 1500
QY 1436 GCTTGAGACTTTCCTTTGTGGTAGGACTTGGAGAGAAATCCCGTGCACCTTCACTAA 1495
Db 1501 GCTTGAGACTTTCCTTTGTGGTAGGACTTGGAGAGAAATCCCGTGCACCTTCACTAA 1560
1496 CCCTCTGACATACCTCCACACCCAGTTGATGCTTTCGTAATAAAGAGATTGGGATT 1554
1561 CCCTCTGACATACCTCCACACCCAGTTGATGCTTTCGTAATAAAGAGATTGGGATT 1619

RESULT 3
AX136115 1631-bp DNA linear PAT 30-MAY-2001
LOCUS
DEFINITION Sequence 37 from Patent EP1067182.
ACCESSION AX136115
VERSION AX136115.1 GI:14272523
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 1631)
Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.
TITLE Secretory protein or membrane protein
JOURNAL Patent: EP 1067182-A 37 10-JAN-2001;
Helix Research Institute (JP)
FEATURES
Location/Qualifiers
source 1..1631
/organism="Homo sapiens"
/db_xref="taxon:9606"
16..1134
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC39700.1"
/db_xref="GI:14272524"
/translation="MAVLAPLIAVYSVPRLSWLAQPYLLSALLSRAFLVLRKLPP
LCHGLPTREDNCPDQWREVEILMFUSAIWMNRRSITVEQHIGNIFMFSKVANT
ILFLRLDIRMGLLYITLCIVFLMTCKPPLYMGPEYIKYFNDRITIDEEERDKRWIV
EFFANWSDCSFAPYADLSLKYNTGTFNGFVDVGRYDSTRYKVSPLTKQLP
TLILFOGKEAMRPOIDKKGRAVSFTSEENVIREFNHLYORAKLSKAGDNIPE
EOPVAOPQCOMGKTRINKILTAVLPLLSIPGSFHNHKEPAAAFYLCFFPGCDW
GQHAASDFKASRELSGLQELPGCQLFHWKSKREIS"
BASE COUNT 411 a 397 c 378 g 445 t
ORIGIN

Query Match 88.1%; Score 1370.2; DB 6; Length 1631;
Best Local Similarity 92.7%; Pred No. 0;
Matches 1507; Conservative 0; Mismatches 3; Indels 115; Gaps 2;

QY 44 GTTAGGCGCGAAAGATGGCGTCTTGGCACCTCTAATTGCCTCGTCTGTTTCGGTGGCG 103
Db 1 GTTAGGCGCGAAAGATGGCGTCTTGGCACCTCTAATTGCCTCGTCTGTTTCGGTGGCG 60
```

Db	1140	AGGGGAAATGGTTTCCCTCCCAAGCTTGGGTGAGTGTAACTGCTTATCAGCTATTCA	1199	
Qy	1130	GACATCTCCATGGTTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTCTTATGTTGACCT	1189	
Db	1200	GACATCTCCATGGTTTCTCCATGAACCTCTGTGGTTTCATCATCTCTTCTTATGTTGACCT	1259	
Qy	1190	GCACAGCTTGGTTAGACCTAGATTTAAACCTAAGCTAAGATGCTGGGTATAGACGCTA	1249	
Db	1260	GCACAGCTTGGTTAGACCTAGATTTAAACCTAAGCTAAGATGCTGGGTATAGACGCTA	1319	
Qy	1250	AGAAATTTCCCCCAAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCTGTTATGTTCTTC	1309	
Db	1320	AGAAATTTCCCCCAAGGACTCTTGCTTCCCTTAAGCCCTTCTGGCTTCTGTTATGTTCTTC	1379	
Qy	1310	ATTTAAAGTAAAGCTTAACCTTGTCTGCTAGTCTTAAGGAGAAACCTTTAAACCAAACT	1369	
J	1380	ATTTAAAGTAAAGCTTAACCTTGTCTGCTAGTCTTAAGGAGAAACCTTTAAACCAAACT	1439	
Qy	1370	TTTTATCATTTGAAGACAATATTGAACAACCCCTATTTTGTGGGATTGAGAGGGGTGA	1429	
Db	1440	TTTTATCATTTGAAGACAATATTGAACAACCCCTATTTTGTGGGATTGAGAGGGGTGA	1499	
Qy	1430	ATAGAGGCTTGAGACTTTCTCTTGTGTGTAGGACTTGGAGGAGAAATCCCTGGACTTT	1489	
Db	1500	ATAGAGGCTTGAGACTTTCTCTTGTGTGTAGGACTTGGAGGAGAAATCCCTGGACTTT	1559	
Qy	1490	CACTAACCTCTGACATACCTCCCAACCCAGCTTGATGGCTTTCGTAATAAAAGATG	1549	
Db	1560	CACTAACCTCTGACATACCTCCCAACCCAGCTTGATGGCTTTCGTAATAAAAGATG	1619	
Qy	1550	GGATT 1554		
Db	1620	GGATT 1624		
RESULT 4				
AX375964				
LOCUS	AX375964	1620 bp	DNA	linear
DEFINITION	Sequence 31 from Patent WO0168848.			PAT 01-MAR-2002
ACCESSION	AX375964			
VERSION	AX375964.1	GI:19170375		
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	Baker,K.P., Chen,J., Desnoyers,L., Goddard,A., Godowski,P.J.,			
	Gurney,A.L., Pan,J., Smith,V., Watanabe,C.K., Wood,W.I. and			
	Zhang,Z.			
TITLE	Secreted and transmembrane polypeptides and nucleic acids encoding			
	the same			
JOURNAL	Patent: WO 0168848-A 31.20-SEP-2001;			
Genentech, Inc. (US)				
FEATURES	Location/Qualifiers			
source	1..1620			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
BASE COUNT	407 a 394 c 373 g 442 t			4 others
ORIGIN				
Query Match	88.1%;	Score 1370;	DB 6;	Length 1620;
Best Local Similarity	92.7%;	Pred. No. 0;		
Matches 1494;	Conservative	0;	Mismatches	4;
			Indels	114;
			Gaps	1;
Qy	57	AGATGGGGCTTGGCACCTTAATGCTCTGCTGATTCGGTCCGCGACTTTCACGAT	116	
Db	1	AGATGGGGCTTGGCACCTTAATGCTCTGCTGATTCGGTCCGCGACTTTCACGAT	60	
Qy	117	GGCTCGGCCAACCTTACTACCTTCTGTGGGCCCTGCTCTGCTGCTTCTCTACCTGTA	176	
Db	61	GGCTCGGCCAACCTTACTACCTTCTGTGGGCCCTGCTCTGCTGCTTCTCTACCTGTA	120	

Qy	177	GGAAACTGCGCGCTCTGCCACGGTCTGCCACCGCAACGCGGTAACCCCGTGTG	236	
Db	121	GGAAACTGCGCGCTCTGCCACGGTCTGCCACCGCAACGCGGTAACCCCGTGTG	180	
Qy	237	ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCATTTGATGATGAAGA	296	
Db	181	ACTTTGACTGGAGAGAGTGGAGATCCTGATGTTTCTCAGTGCATTTGATGATGAAGA	240	
Qy	297	ACGCGAGATGCCA-----	308	
Db	241	ACGCGAGATGCCA-----	300	
Qy	309	-----	308	
Db	301	CCAACAACAATCTTTTCTCCCGCTTGGATATTCGCATGGGCCCTACTTTACATCACA	360	
Qy	309	-----TGTTCTCTGATGAGCTGCAAAACCCCTTATATATGGCCCTGAGTATATCAAGT	362	
Db	361	GCATAGTGTCTCTGATGAGCTGCAAAACCCCTTATATATGGCCCTGAGTATATCAAGT	420	
Qy	363	ACTTCAATGATAAACCAATGATGAGAACTAGAACGGGACAAAGAGGGTCACTTGGATTG	422	
Db	421	ACTTCAATGATAAACCAATGATGAGAACTAGAACGGGACAAAGAGGGTCACTTGGATTG	480	
Qy	423	TGGAGTCTTTTGGCAATTTGGTAACTGCTGCTAACTGCTGCTAACTTTGCCCTATCTAT	482	
Db	481	TGGAGTCTTTTGGCAATTTGGTAACTGCTGCTAACTGCTGCTAACTTTGCCCTATCTAT	540	
Qy	483	TCTCCCTTAAATACAACTGTCACAGGGCTAAATTTTGGAGAGGTGGATTTGGAGCTATA	542	
Db	541	TCTCCCTTAAATACAACTGTCACAGGGCTAAATTTTGGAGAGGTGGATTTGGAGCTATA	600	
Qy	543	CTGATGTTAGTACGCGGTACAAAGTGAAGCACATCACCCCTCACCAAGCACTCCCTACCC	602	
Db	601	CTGATGTTAGTACGCGGTACAAAGTGAAGCACATCACCCCTCACCAAGCACTCCCTACCC	660	
Qy	603	TGATCTGTTCGAAGTGCAGAGGAGCAATGCGGGGCCACAGATTGACAAAGGAC	662	
Db	661	TGATCTGTTCGAAGTGCAGAGGAGCAATGCGGGGCCACAGATTGACAAAGGAC	720	
Qy	663	GGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCGAGAAATTTAACTTAAATGAGC	722	
Db	721	GGCTGTCTCATGGACCTTCTCTGAGGAGAAATGTGATCGAGAAATTTAACTTAAATGAGC	780	
Qy	723	TATACAGCGGGCCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTG	782	
Db	781	TATACAGCGGGCCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAGCTG	840	
Qy	783	TGGCTTCAACCCCGCCAGTGTGATGGGAAACAAAGAGGATTAATAAGATCCTC	842	
Db	841	TGGCTTCAACCCCGCCAGTGTGATGGGAAACAAAGAGGATTAATAAGATCCTC	900	
Qy	843	ACTTTGGCAGTGTCTCTCTCTCAATTCAGGCTCTTCCATACCAACACAGCCTGAG	902	
Db	901	ACTTTGGCAGTGTCTCTCTCTCAATTCAGGCTCTTCCATACCAACACAGCCTGAG	960	
Qy	903	GCTGAGCCTTTTATTTATCTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCT	962	
Db	961	GCTGAGCCTTTTATTTATCTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCT	1020	
Qy	963	TCTGATTTTAAAGAGGATCTAGGGAATTTGTCAGGCACCTTACAGGAAGCCTTGCATGC	1022	
Db	1021	TCTGATTTTAAAGAGGATCTAGGGAATTTGTCAGGCACCTTACAGGAAGCCTTGCATGC	1080	
Qy	1023	TCTGCCCACTTTTCACTGGAGCAGAAAGAGATCTCATAGGACGGAGGGGAAATGTT	1082	
Db	1081	TCTGCCCACTTTTCACTGGAGCAGAAAGAGATCTCATAGGACGGAGGGGAAATGTT	1140	
Qy	1083	TTCCCTTCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATG	1142	
Db	1141	TTCCCTTCAAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATG	1200	

```
QY 1143 TTTCTCCATGAAGTCTGTGGTTTCATCATCTCTTCTAGTGTGACCTGCACAGCTTGTT 1202
DB 1201 TTTCTCCATGAAGTCTGTGGTTTCATCATCTCTTCTAGTGTGACCTGCACAGCTTGTT 1260
QY 1203 AGACCTAGATTTAACCTTAAGGTAAGATGCTGGGGTATAGAACGCTAAGAAATTTCCGCC 1262
DB 1261 AGACCTAGATTTAACCTTAAGGTAAGATGCTGGGGTATAGAACGCTAAGAAATTTCCGCC 1320
QY 1263 AAGGACTCTTGTCTTAAAGCCCTTCTAGCCCTTCTGGCTTCTGGCTTCTTCAATTAAGATATA 1322
DB 1321 AAGGACTCTTGTCTTAAAGCCCTTCTAGCCCTTCTGGCTTCTGGCTTCTTCAATTAAGATATA 1380
QY 1323 GCCTAACTTTTGTCTGCTAGTCTTAAAGGTAAGATGCTGGGGTATAGAACGCTAAGAAATTT 1382
DB 1381 GCCTAACTTTTGTCTGCTAGTCTTAAAGGTAAGATGCTGGGGTATAGAACGCTAAGAAATTT 1440
QY 1383 GACAATATTGAACAACCCCTTATTTTGTGGGGTATAGAACGCTTGAATAGAGCTTTGAG 1442
DB 1441 GACAATATTGAACAACCCCTTATTTTGTGGGGTATAGAACGCTTGAATAGAGCTTTGAG 1500
QY 1443 ACTTCTCTTGTGTGTAGGACTTGGAGGAAATCCCTGGACTTTCACCTAACCCCTCTG 1502
DB 1501 ACTTCTCTTGTGTGTAGGACTTGGAGGAAATCCCTGGACTTTCACCTAACCCCTCTG 1560
Y 1503 ACATACCTCCACACACAGTTGATGCTTTCCGCTTAATAAAAGATTGGGATT 1554
DB 1561 ACATACCTCCACACAGTTGATGCTTTCCGCTTAATAAAAGATTGGGATT 1612
```

```
RESULT 5
AF132965
LOCUS AF132965 1669 bp mRNA linear PRI 18-MAY-2000
DEFINITION Homo sapiens CGI-31 protein mRNA, complete cds.
ACCESSION AF132965
VERSION AF132965.1 GI:4680700
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1669)
AUTHORS Lai, C.H., Chou, C.Y., Ch'ang, L.Y., Liu, C.S. and Lin, W.
TITLE Identification of novel human genes evolutionarily conserved in
Caenorhabditis elegans by comparative proteomics
JOURNAL Genome Res. 10 (5), 703-713 (2000)
MEDLINE 20272150
PUBMED 10810093
REFERENCE 2 (bases 1 to 1669)
AUTHORS Lin, W.-C.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-1999) Institute of Biomedical Sciences, Academia
Sinica, No. 128, Sec. II, Academia Road, Taipei 115, Taiwan
FEATURES
source 1..1669
/organism="Homo sapiens"
/db_xref="taxon.9606"
/chromosome="11"
50..940
/codon_start=1
/product="CGI-31 protein"
/db_xref="AAD2740.1"
/protein_id="A4680701"
/db_xref="GI:4680701"
/translation="MAVLAFLIALVSVPLRWLAQPYLLSALLSAFLVKKLLP
LCHLPTQREDGNPCDFWREVEILMFLSAIVMKNRSTVEQHIGNIEFMFKVANT
ILFRLDIRMLLYITLCIVFLMTCKPPLYMPEYIKYFNDKTIIDELERDKRVTIV
EPFANNSDCOSFAPYVADLSLKYNCTGLNFKGVGVYDVSTRYKVSTVPLTKOLP
TLILFQGGKEAMERPDIDKGRAVSWTFSEENVIRENLNELYQRAKKLSKAGDNIFE
EQPVASTPTVSGENKKDK"
BASE COUNT 420 a 407 c 391 g 451 t
ORIGIN
Query Match 87.2%; Score 1355.4; DB 9; Length 1669;
Best Local Similarity 92.1%; Pred. No. 0;
```

```
Matches 1532; Conservative 0; Mismatches 11; Indels 120; Gaps 5;
QY 10 GGGCGAGAGACTAGCAGCGCGGAGCAGTGGCGTTACGGCCCAAAAGATGGCGGCTTT 69
DB 1 GGGCGAGAGACTAGCAGCGCGGAGCAGTGGCGTTACGGCCCAAAAGATGGCGGCTTT 60
QY 70 GGCACCTCTAATCTCTCGTGTATTCGGTCCCGGACTTTACAGATGGCTCGCCCAACC 129
DB 61 GGCACCTCTAATCTCTCGTGTATTCGGTCCCGGACTTTACAGATGGCTCGCCCAACC 120
QY 130 TTACTACTCTCTGTCGGCCCTGCTCTCTGCTGCTCTCTACTCTGTCGAGAACTGCCGCC 189
DB 121 TTACTACTCTCTGTCGGCCCTGCTCTCTGCTGCTCTCTACTCTGTCGAGAACTGCCGCC 180
QY 190 GCTCTGCACGGCTCTGCCACCAACCGGAAGAGCGGTAAACCGTGTGACTGTGAG 249
DB 181 GCTCTGCACGGCTCTGCCACCAACCGGAAGAGCGGTAAACCGTGTGACTGTGAG 240
QY 250 AGAAGTGGAGATCTGATGTTTCTCAGTGCCATTTGATGATGAAGAACCGCAGATCCA- 308
DB 241 AGAAGTGGAGATCTGATGTTTCTCAGTGCCATTTGATGATGAAGAACCGCAGATCCA 300
QY 309 ----- 308
DB 301 CACTGTGGAGACACATATAGGCAACATTTTCATGTTTAGTAAAGTGGCCAAACAATTTCT 360
QY 309 -----TGTTCCT 315
DB 361 TTTCTCCGCTTGGATATTCGCATGGGCTACTTTACATCACACTCTGCATAGTGTTCCT 420
QY 316 GATCAGCTGCAAAACCCCTTATATATGGGCGCTGAGTATATCAAGTACTTCAATGATAA 375
DB 421 GATCAGCTGCAAAACCCCTTATATATGGGCGCTGAGTATATCAAGTACTTCAATGATAA 480
QY 376 AACCATTTGATGAGGAACCTAGAACGGGACAGAGGCTCATTGGATGTGGAGTTCTTTGC 435
DB 481 AACCATTTGATGAGGAACCTAGAACGGGACAGAGGCTCATTGGATGTGGAGTTCTTTGC 540
QY 436 CAATTTGCTTAATCACTGCCAATCATTTGGCCCTTATCTATGCTGACCTCTCCCTTAATA 495
DB 541 CAATTTGCTTAATCACTGCCAATCATTTGGCCCTTATCTATGCTGACCTCTCCCTTAATA 600
QY 496 CACTGTACAGGGCTAAATTTTGGGAAGTGGATGTTGGAGCGCTATACCTGATGTTAGTAC 555
DB 601 CACTGTACAGGGCTAAATTTTGGGAAGTGGATGTTGGAGCGCTATACCTGATGTTAGTAC 660
QY 556 GCGGTACAAAGTACGACATCACCCCTCACCAAGCACTCCCTACCTGATCTCTGTTCCA 615
DB 661 GCGGTACAAAGTACGACATCACCCCTCACCAAGCACTCCCTACCTGATCTCTGTTCCA 720
QY 616 AGGTGGCAAGAGGCAATGGCGGCGCACAGATTGCAAGAAAGAGCGGGCTGCTCATG 675
DB 721 AGGTGGCAAGAGGCAATGGCGGCGCACAGATTGCAAGAAAGAGCGGGCTGCTCATG 780
QY 676 GACCTTCTCTCAGGAGATGTGATCCGAGATTTAACTTAATAGCTATACCGAGCGGC 735
DB 781 GACCTTCTCTCAGGAGATGTGATCCGAGATTTAACTTAATAGCTATACCGAGCGGC 840
QY 736 CAAGAAACTATCAAAAGCTGGAGCAATATCCCTGAGGAGAGCGCTGTGGCTTCAACCCC 795
DB 841 CAAGAAACTATCAAAAGCTGGAGCAATATCCCTGAGGAGAGCGCTGTGGCTTCAACCCC 900
QY 796 CACCACAGTGTAGATGGGAAAAACAAGAGGATAAATAGATCCTCACTTTGGCAGTGC 855
DB 901 CACCACAGTGTAGATGGGAAAAACAAGAGGATAAATAGATCCTCACTTTGGCAGTGC 960
QY 856 TTCTCTCTCTCAATTCAGGCTCTTCCATACCACAGCCCTGAGGCTGCACCCCTTT 915
DB 961 TTCTCTCTCTCAATTCAGGCTCTTCCATACCACAGCCCTGAGGCTGCACCCCTTT 1018
QY 916 ATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGACGCTTCTGATTTTAAAG 975
DB 1019 ATTTATGTTTTCCCTTTGGCTGTGACTGGGTGGGCGAGCATGACGCTTCTGATTTTAAAG 1078
```

976 AGCATCTAGGGAATGTCAGGCACCTACAGGAAGCCCTCCATGC-TGTGGCCAACTG 1034
 1079 AGCATCTAGGGAATGTCAGGCACCTACAGGAAGCCCTCCATGC-TGTGGCCAACTG 1138
 1035 TTTCACTGGAGC-AAGAAAGAGATCTCATAGGCGGAGGGGAAA--TGGTTTCCCTCCA 1091
 1139 TTTCACTGGAGCAAGAAAGAGATCTCATAGGCGGAGGGGAAAATGGTTTCCCTCCA 1198
 1092 AGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTCCAT 1151
 1199 AGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTCCAT 1258
 1152 GAACTCTGTTGTTTATCATCTCTTCTAGTTCAGCTGACAGCTGGTTAGACCTAGA 1211
 1259 GAACTCTGTTGTTTATCATCTCTTCTAGTTCAGCTGACAGCTGGTTAGACCTAGA 1318
 1212 TTTAAACCTTAAGTAACTGCTGGGGTATAGACGCTTAAGAAATTTCCCCCAAGGACTCT 1271
 1319 TTTAAACCTTAAGTAACTGCTGGGGTATAGACGCTTAAGAAATTTCCCCCAAGGACTCT 1378
 1272 TGTCTTCTTAAGCCCTCTCGCTTCGTTTATGCTTCATTAAGTAACTTAAGCCTAAGT 1331
 1379 TGTCTTCTTAAGCCCTCTCGCTTCGTTTATGCTTCATTAAGTAACTTAAGCCTAAGT 1438
 1332 TGTCTGCTAGTCCCTAAGGAGAACTTTTAAACACAACTTTTATCATTTGAAGCAATATT 1391
 1439 TGTCTGCTAGTCCCTAAGGAGAACTTTTAAACACAACTTTTATCATTTGAAGCAATATT 1498
 1392 GAACACCCCTCTATTTGTGGGATTCAGAGGGGTGAATAGAGCTTTGAGACTTTCTCT 1451
 1499 GAACACCCCTCTATTTGTGGGATTCAGAGGGGTGAATAGAGCTTTGAGACTTTCTCT 1558
 1452 TGTGTGTAGGACTTGAGGAGAAATCCCTGGACTTTCATTAACCTCTGACATACCTC 1511
 1559 TGTGTGTAGGACTTGAGGAGAAATCCCTGGACTTTCATTAACCTCTGACATACCTC 1618
 1512 CCACACCCCTGATGCTTCGTTTAAAGAAATTTGGGATT 1554
 1619 CCACACCCCTGATGCTTCGTTTAAAGAAATTTGGGATT 1661

RESULT 6

AL138752/C
 INITIATION
 Human DNA sequence from clone RP11-3J10 on chromosome 9p12-13.3
 Contains part of a novel gene, the 3' end of the SHB (SHB adaptor protein (a Src homology 2) protein) gene, the gene for CG7943 protein, a pseudogene similar to CGI-31 protein, a pseudogene similar to PAICS (phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase), part of a gene for a ribosomal protein, the 3' end of gene for apoptosis-related protein PNAS-3, ESTs, GSSs and CpG islands, complete sequence.
 AL138752
 AL138752.5 GI:8452480
 HTG.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 189271)
 Laired,G.
 Direct Submission
 Submitted (08-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 10, 2000 this sequence version replaced gi:8250065.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>
 This sequence is the entire insert of clone RP11-3J10 This sequence has been finished according to sequence map criteria as follows.
 An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RP11-3J10 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6

FEATURES

source

Location/Qualifiers

1. .189271
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="p12-13.3"
 /clone="RP11-3J10"
 /clone_lib="RPC1-11.1"
 126. .493
 /note="L2 repeat: matches 2325. .2705 of consensus"
 1456. .1537
 /note="MER66B repeat: matches 271. .349 of consensus
 MER66A repeat: matches 263. .341 of consensus"
 2434. .2532
 /note="MIR repeat: matches 23. .122 of consensus"
 3695. .3917
 /note="MIR repeat: matches 9. .256 of consensus"
 4195. .4263
 /note="MIR repeat: matches 83. .152 of consensus"
 5058. .5122
 /note="MIR repeat: matches 39. .103 of consensus"
 5751. .5854
 /note="MIR repeat: matches 65. .169 of consensus"
 7967. .8173
 /note="L2 repeat: matches 2457. .2687 of consensus"
 8421. .8481
 /note="U7 repeat: matches 1. .62 of consensus
 U7 repeat: matches 1. .62 of consensus"
 8638. .8687
 /note="L2 repeat: matches 2645. .2696 of consensus"
 8948. .9233
 /note="AlusX repeat: matches 36. .308 of consensus"
 9410. .9571
 /note="MIR repeat: matches 5. .190 of consensus"
 10596. .10867
 /note="AluJ repeat: matches 8. .276 of consensus"
 11536. .11593
 /note="MER5A repeat: matches 128. .186 of consensus"
 11937. .11971
 /note="MIR repeat: matches 112. .146 of consensus"
 12275. .12551
 /note="AluJ repeat: matches 1. .274 of consensus"
 12555. .12885
 /note="L1MA9 repeat: matches 5880. .6216 of consensus
 L1MA9 repeat: matches 5880. .6216 of consensus"
 12886. .13189
 /note="AluJ repeat: matches 1. .313 of consensus"
 13190. .14503
 /note="L1MA9 repeat: matches 4579. .5880 of consensus
 L1MA9 repeat: matches 4579. .5880 of consensus"
 14504. .15045
 /gene="BA3J10.1"
 <14504. .15045
 gene
 mRNA

QY 964 CTGATTTTAAAGAGCATCTAGGAATTTGTCAGGCACCTCAGAGAAAGCCCTGCATGCT 1023
 |||
 Db 58670 CTGATTTTAAAGAGCATCTAGGAATTTGTCAGGCACCTCAGAGAAAGCCCTGCATGCT 58611
 |||
 QY 1024 GTGGCCAACTGTTTCACTGGAGCAAGAAGAGATCTCTATAGGAGGGGGAATGGTT 1083
 |||
 Db 58610 GTGGCCAACTGTTTCACTGGAGCAAGAAGAGATCTCTATAGGAGGGGGAATGGTT 58551
 |||
 QY 1084 TCCTTCAAGCTTGGGTCAGTGTGTTACTGCTTATCAGCTATTCAGACATCTCCATGCT 1143
 |||
 Db 58550 TCCTTCAAGCTTGGGTCAGTGTGTTACTGCTTATCAGCTATTCAGACATCTCCATGCT 58491
 |||
 QY 1144 TTCTCCATGAACCTCTGTGTTTTCATCTTCTTGTACCTGCACAGCTTGGTTA 1203
 |||
 Db 58490 TTCTCCATGAACCTCTGTGTTTTCATCTTCTTGTACCTGCACAGCTTGGTTA 58431
 |||
 - 1204 GACCTAGATTAAACCTTAAGTAAGATGCTGGGGTATAGACGCTTAAGAAATTTCCCCCA 1263
 |||
 Db 58430 GACCTAGATTAAACCTTAAGTAAGATGCTGGGGTATAGACGCTTAAGAAATTTCCCCCA 58371
 |||
 QY 1264 AGACCTCTGCTTCCCTTAAGCCCTCTGGCTTCGTTTATGCTTCATTAAGATTAAG 1323
 |||
 Db 58370 AGACCTCTGCTTCCCTTAAGCCCTCTGGCTTCGTTTATGCTTCATTAAGATTAAG 58311
 |||
 QY 1324 CCTAACT 1330
 |||
 Db 58310 CCTGGCT 58304

RESULT 7
 AC073598
 LOCUS
 DEFINITION Homo sapiens chromosome 10 clone CTC-268N23, WORKING DRAFT
 SEQUENCE, 9 unordered pieces.
 AC073598
 VERSION AC073598.3 GI:20279338
 KEYWORDS HTG, HTGS, PHASE1, HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 112389)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayalew,M., Banks,T., Barbara,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Burch,C., Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mitchell,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,

Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 112389)
 Worley,K.C.
 Direct Submission
 Submitted (26-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 112389)
 Worley,K.C.
 Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Apr 24, 2002 this sequence version replaced gi:16118028.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

Project Information
 Center project name: HMRT
 Center clone name: CTC-268N23

Summary Statistics
 Sequencing vector: M13

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 104470 bases at least Q40

Consensus quality: 107539 bases at least Q30

Consensus quality: 109442 bases at least Q20

Estimated insert size: 111011; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 9 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

2191: contig of 2191 bp in length

2192: gap of unknown length

2292: contig of 3607 bp in length

5998: gap of unknown length

5999: contig of 5408 bp in length

11407: gap of unknown length

11507: contig of 10233 bp in length

21740: gap of unknown length

21840: contig of 10117 bp in length

31957: gap of unknown length

32057: contig of 14718 bp in length

46775: gap of unknown length

46875: contig of 21344 bp in length

68219: gap of unknown length

68319: contig of 17636 bp in length

85955: gap of unknown length

86055: contig of 26335 bp in length.

Location/Qualifiers

1..112389

/organism="Homo sapiens"

FEATURES
 source


```
/db_xref="taxon:9606"
/chromosome="10"
/clon="CTC-268N23"
BASE COUNT 29689 a 25997 c 25893 g 29967 t 843 others
ORIGIN

Query Match 60.1%; Score 934.2; DB 2; Length 112389;
Best Local Similarity 95.8%; Pred. No. 1.9e-281;
Matches 983; Conservative 0; Mismatches 39; Indels 6; Gaps 2;

QY 309 TGTCTCTGATGACGTGCAAAACCC-----CCCTATATATGGCCCTGAGTATATCAAGTA 363
DB 19739 TGTCTCTGATGACGTGCAAAACCC-----CCCTATATATGGCCCTGAGTATATCAAGTA 19798
QY 364 CTTCAATGATAAAACCACTTATGAGGAAGTATAGAACGGGACAGAGGGTCACTTGGATTGT 423
DB 19799 CTTCAATGATAAAACCACTTATGAGGAAGTATAGAACGGGACAGAGGGTCACTTGGATTGT 19858
QY 424 GGAGTTCTTTGCCAATTTGCTTAATGACTGCGCAATCATTTTGGCCCTATCTATCTGACCT 483
DB 19859 GGAGTTCTTTGCCAATTTGCTTAATGACTGCGCAATCATTTTGGCCCTATCTATCTGACCT 19918
QY 484 CTCCTCTTAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGAGCGCTATAC 543
DB 19919 CTCCTCTTAATACAACTGTACAGGGCTAAATTTTGGGAAGGTGGATGTTGGAGCGCTATAC 19978
QY 544 TGATGTTAGTACGGGTACAAAGTGTAGACATCACCCCTCACCAAGCAACTCCCTACCT 603
DB 19979 TGATGTTAGTACGGGTACAAAGTGTAGACATCACCCCTCACCAAGCAACTCCCTACCT 20038
QY 604 GATCTGTTCCAGGTGGCAAGGAGCAATTCGGCGGCCACACATTTGACAAAGAGGACG 663
DB 20039 GATCTGTTCCAGGTGGCAAGGAGCAATTCGGCGGCCACACATTTGACAAAGAGGACG 20098
QY 664 GGTGCTCTCATGGACC-TTCTCTGAGGAGAAATGTGATCCGAGAAATTAACCTTAATGAGC 722
DB 20099 GGTGCTCTCATGGACC-TTCTCTGAGGAGAAATGTGATCCGAGAAATTAACCTTAATGAGC 20158
QY 723 TATACAGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGAGCGCTG 782
DB 20159 TATACAGCGGGCCAAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGAGCGCTG 20218
QY 783 TGGCTTCAACCCCAACAGTGTACAGTGGGGAACAAAGAGGATAAATAGATCCTC 842
DB 20219 TGGCTTCAACCCCAACAGTGTACAGTGGGGAACAAAGAGGATAAATAGATCCTC 20278
QY 843 ACTTTGGCAGTGTCTCTCTCTCAATTCAGGCTCTTTCCATACCAACAGCGCTGAG 902
DB 20279 ACTTTGGCAGTGTCTCTCTCTCAATTCAGGCTCTTTCCATACCAACAGCGCTGAG 20338
QY 903 GGTGAGCGCTTTTATTTATTTTCCCTTTGGCTGTGACTGGTGGGGGAGCATGACGT 962
DB 20339 ACTAGAGCGCTTTTATTTATTTTCCCTTTGGCTGTGACTGGTGGGGGAGCATGACGT 20398
QY 963 TCTGATTTTAAAGAGCATCTAGGGAATTTCTAGGCAACCTACAGAGGCGCTGCCATGC 1022
DB 20399 TCTGATTTTAAAGAGCATCTAGGGAATTTCTAGGCAACCTACAGAGGCGCTGCCATGC 20458
QY 1023 TGTGGCAACTGTGTTCACTGGCAAGAAAGAGATCTCATAGGAGGAGGGGGAATPGT 1082
DB 20459 TGTGGCAACTGTGTTCACTGGCAAGAAAGAGATCTCATAGGAGGAGGGGGAATPGT 20518
QY 1083 TTCCCTCCAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 1142
DB 20519 TTCCCTCCAGCTTGGGTGAGTGTGTTAACTGCTTATCAGCTATTCAGACATCTCCATGG 20578
QY 1143 TTTCTCCATGAACCTCTGGTTTTCATCATCTCTTCTAGTGTACCTGACAGCTTGGTT 1202
DB 20579 TTTCTCCATGAACCTCTGGTTTTCATCATCTCTTCTAGTGTACCTGACAGCTTGGTT 20638
QY 1203 AGACCTTAGATTTAAACCTTAAGTAAAGTGTGCGGTATAGAACGCTAAAGAAATTTTCCCC 1262
DB 20639 AGACCTTAGATTTAAACCTTAAGTAAAGTGTGCGGTATAGAACGCTAAAGAAATTTTCCCC 20698
```

```
1263 AAGGACTTGTCTTCCCTTAAGCCCTTCTGCTCGTTTATGCTTCTTCAATTAAGATATAA 1322
12699 AAGGACTTGTCTTCCCTTAAGCCCTTCTGCTCGTTTATGCTTCTTCAATTAAGATATAA 20758
1323 GCCTAAGT 1330
20759 GCCTGAGT 20766

RESULT 8
AP001931
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:RP11-691N7, complete
DEFINITION sequences.
ACCESSION AP001931
VERSION AP001931.5 GI:21218128
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-691N7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 110116)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On May 27, 2002 this sequence version replaced gi:20334325.
FEATURES
Location/Qualifiers
1..110116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-691N7"
BASE COUNT 29082 a 23179 c 24671 g 33184 t
ORIGIN

Query Match 55.8%; Score 868; DB 9; Length 110116;
Best Local Similarity 100.0%; Pred. No. 1.2e-260; Indels 0; Gaps 0;
Matches 868; Conservative 0; Mismatches 0;

QY 687 AGGAGATGTGATCCGAGAAATTTAACTTAATGAGCTATACAGCGGGCCCAAGAACTAT 746
DB 20154 AGGAGATGTGATCCGAGAAATTTAACTTAATGAGCTATACAGCGGGCCCAAGAACTAT 20213
QY 747 CAAGAGCTGAGACAATATCCCTGAGGAGAGCGCTGTGGCTTCAACCCCAACACAGTGT 806
DB 20214 CAAGAGCTGAGACAATATCCCTGAGGAGAGCGCTGTGGCTTCAACCCCAACACAGTGT 20273
QY 807 CAGATGGGAAAACAAGAGATAAATAGATCTCACTTTGGCAGTGTCTCTCTCTG 866
DB 20274 CAGATGGGAAAACAAGAGATAAATAGATCTCACTTTGGCAGTGTCTCTCTCTG 20333
QY 867 TCAATTCAGGCTCTTTCCATAACCAAGCCTGAGCTGCAGCTTTTATTTATGTTTT 926
DB 20334 TCAATTCAGGCTCTTTCCATAACCAAGCCTGAGCTGCAGCTTTTATTTATGTTTT 20393
QY 927 CCCTTTGGCTGTGACTGGGTGGGGAGCATGCAAGCTTCTGATTTTAAAGAGCATCTAGG 986
DB 20394 CCCTTTGGCTGTGACTGGGTGGGGAGCATGCAAGCTTCTGATTTTAAAGAGCATCTAGG 20453
QY 987 GAATTTGTCAGGACCCCTACAGGAGGCTGCCATGCTGTGGCAACTGTTTCACTGAGC 1046
```

```

Db      20454 GAATTGTCAGCACCTACAGGAAGCCCTGCCATGCTGTGGCCAACTGTTTCACTGGAGC 20513
QY      1047 AAGAAGAGATCTCATAGACGAGGAGGGAATGTTTCCCTCCAAGCTTGGGTCACTGT 1106
Db      20514 AGAAGAGATCTCATAGACGAGGAGGGAATGTTTCCCTCCAAGCTTGGGTCACTGT 20573
QY      1107 GTTAACCTGTTATCAGCTATTGAGACATCTCCATGTTTCTCCATGAAACTCTGTGGTTT 1166
Db      20574 GTTAACCTGTTATCAGCTATTGAGACATCTCCATGTTTCTCCATGAAACTCTGTGGTTT 20633
QY      1167 CATCATCTCTTCTAGTTGACCTGCACACGCTTGGTTAGACCTAGATTAAACCTAAGGTA 1226
Db      20634 CATCATCTCTTCTAGTTGACCTGCACACGCTTGGTTAGACCTAGATTAAACCTAAGGTA 20693
QY      1227 AGATCTGGGGTATAGAACCTTAAGAAATTTCCCCCAAGGACTCTTCTTCTTAAAGCCC 1286
Db      20694 AGATCTGGGGTATAGAACCTTAAGAAATTTCCCCCAAGGACTCTTCTTCTTAAAGCCC 20753
QY      1287 TCTCGCTTCGTTTATGCTCTTCAATAAAGTATAGCCCTAATTTGCTGCTAGTCTCTAA 1346
Db      20754 TCTCGCTTCGTTTATGCTCTTCAATAAAGTATAGCCCTAATTTGCTGCTAGTCTCTAA 20813
QY      1347 GGAGAAACCTTTAACCACCAAGTTTATCATTTGAGACAATATGAAACAACCCCTATT 1406
Db      20814 GGAGAAACCTTTAACCACCAAGTTTATCATTTGAGACAATATGAAACAACCCCTATT 20873
QY      1407 TTGTGGGATTCAGAGGGGTCAATAGAGGCTTGAGACTTTCTTGTGTGGTAGACATT 1466
Db      20874 TTGTGGGATTCAGAGGGGTCAATAGAGGCTTGAGACTTTCTTGTGTGGTAGACATT 20933
QY      1467 GGAGAGAAATCCCTGGACTTTCACTAACCCCTCTGACATCTCCCAACCCAGTTGAT 1526
Db      20934 GGAGAGAAATCCCTGGACTTTCACTAACCCCTCTGACATCTCCCAACCCAGTTGAT 20993
QY      1527 GCCTTCGTAATAAAGATTGGGATT 1554
Db      20994 GCCTTCGTAATAAAGATTGGGATT 21021

RESULT 9
AP001082
LOCUS      130642 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone CMB9-77p23 map 11q12, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
'CESSION
ASTON
KEYWORDS
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 130642)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 130,642 genomic DNA of 11q12
Published Only in Database (2000)
2 (bases 1 to 130642)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 130,642 genomic DNA of 11q12
Submitted (25-JAN-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail:hattori@psc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:6997796.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
----- Project Information
-----

```

```

Center project name: HumDraft11
Center clone name: CMB9-77p23
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 109722 bases at least Q40
Consensus quality: 118358 bases at least Q30
Consensus quality: 123353 bases at least Q20
Insert size: 127542; sum-of-contigs
Quality coverage: 4.10x in Q20 bases; sum-of-contigs
-----

```

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 14868 contig of 14868 bp in length
14969 24185 contig of 9217 bp in length
24286 32497 contig of 8212 bp in length
32598 34006 contig of 6809 bp in length
39507 44898 contig of 5392 bp in length
44999 50122 contig of 5124 bp in length
50223 57989 contig of 7767 bp in length
58090 64088 contig of 5999 bp in length
68556 68556 contig of 4368 bp in length
72259 72259 contig of 3603 bp in length
76692 76692 contig of 4333 bp in length
80690 80690 contig of 3898 bp in length
80791 84253 contig of 3463 bp in length
84354 88683 contig of 4330 bp in length
91721 91721 contig of 2938 bp in length
95301 95301 contig of 3480 bp in length
99413 99413 contig of 4012 bp in length
99514 103118 contig of 3605 bp in length
103219 106147 contig of 2929 bp in length
106248 109453 contig of 3206 bp in length
109554 111611 contig of 2058 bp in length
111712 113460 contig of 1749 bp in length
113561 115886 contig of 2326 bp in length
115987 118771 contig of 2785 bp in length
118872 120623 contig of 1752 bp in length
120724 122472 contig of 1749 bp in length
122573 123753 contig of 1181 bp in length
123854 124856 contig of 1003 bp in length
124957 126517 contig of 1561 bp in length
126618 128115 contig of 1498 bp in length
128216 129271 contig of 1056 bp in length
129372 130642 contig of 1271 bp in length

Sequence updated (02-FEB-2000)
Sequence updated (26-MAY-2000)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 14868: contig of 14868 bp in length
* 14869 14968: gap of 100 bp
* 14969 24185: contig of 9217 bp in length
* 24186 24285: gap of 100 bp
* 32497 32497: contig of 8212 bp in length
* 32498 32597: gap of 100 bp
* 32598 34006: contig of 6809 bp in length
* 34007 39506: gap of 100 bp
* 39507 44898: contig of 5392 bp in length
* 44899 44998: gap of 100 bp
* 44999 50122: contig of 5124 bp in length
* 50123 50222: gap of 100 bp

```



```

Db      52555 GGAGAAACCTTTAACCAACAAAGTTTATCATTTGAAGACAATATTGAACACCCCTATT 52614
QY      1407 TTGTGGGATTGAGAGGGGTGATAGAGGCTTGAGACTTTCCTTGTGTGTAGACTT 1466
        |||||||
Db      52615 TTGTGGGATTGAGAGGGGTGATAGAGGCTTGAGACTTTCCTTGTGTGTAGACTT 52674
        |||||||
QY      1467 GGAGGAGAAATCCCTCGACTTTCACATAACCTCTGACATACCTCCACACCCAGTTGAT 1526
        |||||||
Db      52675 GGAGGAGAAATCCCTCGACTTTCACATAACCTCTGACATACCTCCACACCCAGTTGAT 52734
        |||||||
QY      1527 GGCTTCCGTAATAAAAGATTGGGATT 1554
        |||||||
Db      52735 GGCTTCCGTAATAAAAGATTGGGATT 52762
        |||||||

>RESULT 10
309438
>JUS
DEFINITION      AC009438      158349 bp      DNA      linear      HTG 04-SEP-2000
SEQUENCE, 25 unordered pieces.
ACCESSION      AC009438
VERSION        AC009438.3 GI:9966254
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 158349)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
                Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
                Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A.,
                Cooke,P., DeArellano,K., Delapayre,E., Devon,K., Dewar,K.,
                Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
                Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
                Hagos,B., Heaford,A., Horton,L., Howland,J., Jones,C., Kann,L.,
                Karakas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
                Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
                Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
                Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
                Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
                Stange-Rhmann,N., Stojanovic,N., Stone,C., Subramanian,A.,
                Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
                Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
                Direct Submission
                Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                On Sep 4, 2000 this sequence version replaced gi:7655447.
                All repeats were identified using RepeatMasker:
                Smit, A.F.A. & Green, P. (1996-1997)
                http://ftp.genome.washington.edu/RM/RepeatMasker.html
                ----- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www.seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
                ----- Project Information
                Center project name: L1359
                Center clone name: 77_M_17
                ----- Summary Statistics
                Sequencing vector: M13; M7815; 100% of reads
                Chemistry: Dye-terminator-amersham; 5% of reads
                Chemistry: Dye-terminator Big Dye; 95% of reads
                Assembly program: Phrap; version 0.960731.
                Consensus quality: 138242 bases at least Q40
                Consensus quality: 145728 bases at least Q30
                Consensus quality: 149650 bases at least Q20
                Insert size: 172000; agarose-fp
                Insert size: 155949; sum-of-ctnigs
                Quality coverage: 4.0 in Q20 bases; agarose-fp

```

* Quality coverage: 4.4 in Q20 base.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1
 * 3651 3750: contig of 3650 bp in length
 * 3751 4867: contig of 1117 bp in length
 * 4868 4967: gap of 100 bp
 * 4968 6757: contig of 1790 bp in length
 * 6758 6857: gap of 100 bp
 * 6858 7991: contig of 1134 bp in length
 * 7992 8091: gap of 100 bp
 * 8092 9216: contig of 1125 bp in length
 * 9217 9316: gap of 100 bp
 * 9317 10384: contig of 1068 bp in length
 * 10385 10484: gap of 100 bp
 * 10485 11835: contig of 1351 bp in length
 * 11836 11935: gap of 100 bp
 * 11936 13429: contig of 1494 bp in length
 * 13430 13529: gap of 100 bp
 * 13530 15092: contig of 1563 bp in length
 * 15093 15192: gap of 100 bp
 * 15193 17380: contig of 2188 bp in length
 * 17381 17480: gap of 100 bp
 * 17481 20586: contig of 3106 bp in length
 * 20587 20686: gap of 100 bp
 * 20687 24958: contig of 4272 bp in length
 * 24959 25058: gap of 100 bp
 * 25059 30271: contig of 5213 bp in length
 * 30272 30371: gap of 100 bp
 * 30372 35676: contig of 5305 bp in length
 * 35677 35776: gap of 100 bp
 * 35777 40640: contig of 4864 bp in length
 * 40641 40740: gap of 100 bp
 * 40741 46368: contig of 5628 bp in length
 * 46369 46468: gap of 100 bp
 * 46469 53373: contig of 6905 bp in length
 * 53374 53473: gap of 100 bp
 * 53474 60571: contig of 7098 bp in length
 * 60572 60671: gap of 100 bp
 * 60672 67052: contig of 6381 bp in length
 * 67053 67152: gap of 100 bp
 * 67153 75831: contig of 8679 bp in length
 * 75832 75931: gap of 100 bp
 * 75932 84118: contig of 8187 bp in length
 * 84119 84218: gap of 100 bp
 * 84219 95075: contig of 10857 bp in length
 * 95076 95175: gap of 100 bp
 * 95176 110669: contig of 15494 bp in length
 * 110670 110769: gap of 100 bp
 * 110770 139814: contig of 29045 bp in length
 * 139815 139914: gap of 100 bp
 * 139915 158349: contig of 18435 bp in length.

FEATURES

Location/Qualifiers
 1. 158349
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-77M17"
 /clone.lib="RPC1-11 Human Male BAC"
 1. 3650
 /note="assembly_fragment"
 clone_end:sp6
 vector_side:left
 3751. 4867
 /note="assembly_fragment"
 4968. 6757
 misc_feature
 misc_feature
 misc_feature

```

/misc_feature      /note="assembly_fragment"
6858. 7991
/misc_feature      /note="assembly_fragment"
8092. 9216
/misc_feature      /note="assembly_fragment"
9317. 10384
/misc_feature      /note="assembly_fragment"
10485. 11835
/misc_feature      /note="assembly_fragment"
11936. 13429
/misc_feature      /note="assembly_fragment"
13530. 15092
/misc_feature      /note="assembly_fragment"
15193. 17380
/misc_feature      /note="assembly_fragment"
17481. 20586
/misc_feature      /note="assembly_fragment"
20687. 24958
/misc_feature      /note="assembly_fragment"
25059. 30271
/misc_feature      /note="assembly_fragment"
30372. 35676
/misc_feature      /note="assembly_fragment"
35777. 40640
/misc_feature      /note="assembly_fragment"
40741. 46368
/misc_feature      /note="assembly_fragment"
46469. 53373
/misc_feature      /note="assembly_fragment"
53474. 60571
/misc_feature      /note="assembly_fragment"
60672. 67052
/misc_feature      /note="assembly_fragment"
67153. 75831
/misc_feature      /note="assembly_fragment"
75932. 84118
/misc_feature      /note="assembly_fragment"
84219. 95075
/misc_feature      /note="assembly_fragment"
95176. 110669
/misc_feature      /note="assembly_fragment"
110770. 139814
/misc_feature      /note="assembly_fragment"
139915. 158349
/misc_feature      /note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT      40315 a 36712 c 36032 g 42881 t 2409 others
ORIGIN
Query Match      55.8%; Score 868; DB 2; Length 158349;
Best Local Similarity 100.0%; Pred. No. 1.3e-260;
atches 868; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 687 AGGAGATGTGATCCGAGATTTAACTTAATGAGCTATACACGGGGCCCAAGAACTAT 746
DB 144815 AGGAGATGTGATCCGAGATTTAACTTAATGAGCTATACACGGGGCCCAAGAACTAT 144874
QY 747 CAAGAGCTGGAGCAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGT 806
DB 144875 CAAGAGCTGGAGCAATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGT 144934
QY 807 CAGATGGGAAACAAAGAGGATAAATAGATCCTCAGCTTTGGCAGTGTCTCTCCTG 866
DB 144935 CAGATGGGAAACAAAGAGGATAAATAGATCCTCAGCTTTGGCAGTGTCTCTCCTG 144994
QY 867 TCAATTCAGGCTCTTCCATACACAGGCTGAGGCTGCAGCCCTTTATTTATGTTTT 926
DB 144995 TCAATTCAGGCTCTTCCATACACAGGCTGAGGCTGCAGCCCTTTATTTATGTTTT 145054
QY 927 CCCTTGGCTGTGACTGGGTGGGGCAGCATGCAGGCTTCTGATTTAAAGAGGCATCTAGG 986
DB 145055 CCCTTGGCTGTGACTGGGTGGGGCAGCATGCAGGCTTCTGATTTAAAGAGGCATCTAGG 145114

```

```

QY 987 GAATTCCTCAGGCACCTACAGGAAGCCCTGCCATGCTGTGGCCAACTGTTTCACTGAGC 1046
DB 145115 GAATTCCTCAGGCACCTACAGGAAGCCCTGCCATGCTGTGGCCAACTGTTTCACTGAGC 145174
QY 1047 AAGAAAGAGATCTCATAGGACGAGGGGAAATGGTTTCCTCCAAAGCTTGGGTCACTGT 1106
DB 145175 AAGAAAGAGATCTCATAGGACGAGGGGAAATGGTTTCCTCCAAAGCTTGGGTCACTGT 145234
QY 1107 GTTAACCTGCTATCAGCTATTCAGACATCTCCATGGTTTCTCCATGAACCTCTGTGTTT 1166
DB 145235 GTTAACCTGCTATCAGCTATTCAGACATCTCCATGGTTTCTCCATGAACCTCTGTGTTT 145294
QY 1167 CATCATTCCTCTTACTTGACCTGCACAGCTTGGTTAGACCTAGATTTAACCCCTAAGSTA 1226
DB 145295 CATCATTCCTCTTACTTGACCTGCACAGCTTGGTTAGACCTAGATTTAACCCCTAAGSTA 145354
QY 1227 AGATGCTGGGTATAGAACGCTAAGAAATTTTCCCCCAAGGACTCTTCTTCTTAAAGCCC 1286
DB 145355 AGATGCTGGGTATAGAACGCTAAGAAATTTTCCCCCAAGGACTCTTCTTCTTAAAGCCC 145414
QY 1287 TTCTGGCTTCGTTTATGGTCTTCATTAAGATATAAGCCTAACTTTTGCCTAGTCTCTAA 1346
DB 145415 TTCTGGCTTCGTTTATGGTCTTCATTAAGATATAAGCCTAACTTTTGCCTAGTCTCTAA 145474
QY 1347 GGAGAAACCTTTAACCACAAAGTTTTTATCATTTGAAGACAAATATTGAACAAACCCCTATT 1406
DB 145475 GGAGAAACCTTTAACCACAAAGTTTTTATCATTTGAAGACAAATATTGAACAAACCCCTATT 145534
QY 1407 TTCTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAAGACTTTCTCTTGTGTAGGACTTT 1466
DB 145535 TTCTGGGGATTGAGAAGGGGTGAATAGAGGCTTGAAGACTTTCTCTTGTGTAGGACTTT 145594
QY 1467 GGAGGAGAAATCCCTGGACTTTTCACTAACCTCTGACATACCTCCACACACCCAGTTGAT 1526
DB 145595 GGAGGAGAAATCCCTGGACTTTTCACTAACCTCTGACATACCTCCACACACCCAGTTGAT 145654
QY 1527 GCCTTTCCGTAAATAAAGATTTGGGATT 1554
DB 145655 GCCTTTCCGTAAATAAAGATTTGGGATT 145682

RESULT 11
AC090218/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-691N7 map 11, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC090218.3 GI:14718351
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 161397)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 11, clone RP11-691N7
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 161397)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barnes,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Cammarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karakas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,

```



```

Db      36303  CCCTTTGGCTGTGACTGGCTGGGGGACGATGCGAGCTTCTGATTTTAAAGAGGCATCTAGG 36244
QY      987    GAATTCCTAGGACCCCTACAGGAGCCCTCCCATGCTGTGGCCAACTGTTTCACCTGAGC 1046
Db      36243  GAATTCCTAGGACCCCTACAGGAGCCCTCCCATGCTGTGGCCAACTGTTTCACCTGAGC 36184
QY      1047  AAGAAGAGATCTCATAGGACGGAGGGGAAATGGTTTCCCTCCCAAGCTTGGGTCAAGTGT 1106
Db      36183  AAGAAGAGATCTCATAGGACGGAGGGGAAATGGTTTCCCTCCCAAGCTTGGGTCAAGTGT 36124
QY      1107  GTTAAGTCTTATCAGCTATTCAGACATCTCCATGTTTCTCATGAACTCTGTGGTTT 1166
Db      36123  GTTAAGTCTTATCAGCTATTCAGACATCTCCATGTTTCTCATGAACTCTGTGGTTT 36064
QY      1167  CATCATCTCTTCTTAGTTCACCTGCACAGCTTGGTTAGACCTAGATTTAACCCCTAAGGTA 1226
Db      36063  CATCATCTCTTCTTAGTTCACCTGCACAGCTTGGTTAGACCTAGATTTAACCCCTAAGGTA 36004
QY      1227  AGATGCTGGGGTATAGAACGCTAAGAAATTTTCCCAAGGACTTCTGCTTCCCTTAAGCCC 1286
Db      36003  AGATGCTGGGGTATAGAACGCTAAGAAATTTTCCCAAGGACTTCTGCTTCCCTTAAGCCC 35944
QY      1287  TTCTGGCTTCGTTTATGGTCTTCATTAAGTATAAGCCCTAACTTTGTCGCTAGTCTCTAA 1346
Db      35943  TTCTGGCTTCGTTTATGGTCTTCATTAAGTATAAGCCCTAACTTTGTCGCTAGTCTCTAA 35884
QY      1347  GGAGAAACCTTTAACCCACAAGTTTTTATCATTTGAAGACATATTGAACAAACCCCTATT 1406
Db      35883  GGAGAAACCTTTAACCCACAAGTTTTTATCATTTGAAGACATATTGAACAAACCCCTATT 35824
QY      1407  TTGTGGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCCTTTGTGTGGTAGGACTT 1466
Db      35823  TTGTGGGGATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCCTTTGTGTGGTAGGACTT 35764
QY      1467  GGAGGAGAAATCCCTGGACTTTCATTAACCCCTGACATCTCCACACACCCAGCTGTAT 1526
Db      35763  GGAGGAGAAATCCCTGGACTTTCATTAACCCCTGACATCTCCACACACCCAGCTGTAT 35704
QY      1527  GCCTTTCCCTATAAAGATTTGGGATT 1554
Db      35703  GCCTTTCCCTATAAAGATTTGGGATT 35676

RESULT 12
AC092707
LOCUS   AC092707
DEFINITION Homo sapiens chromosome 11 clone RP11-679G21 map 11, WORKING DRAFT
SEQUENCE, 9 unordered pieces.
ACCESSION AC092707
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE  Homo sapiens.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 170540)
    Birren.B., Linton.L., Nusbaum.C., Lander.E., Allen.N.,
    Anderson.S., Barna.N., Bastien.V., Boguslavskiy.L., Boukhgalter.B.,
    Brown.A., Camarata.J., Campopiano.A., Chang.J., Chazaro.B.,
    Choquel.Y., Colangelo.M., Collins.S., Collamore.A., Cook.A.,
    Cooke.P., DeArrellano.K., Dewar.K., Diaz.J.S., Dodge.S., Faro.S.,
    Ferreira.P., Fitzhugh.W., Gage.D., Galagan.J., Gardyna.S.,
    Ginde.S., Gord.S., Goyette.M., Graham.L., Grand-Pierre.N.,
    Hagos.B., Heaford.A., Horton.L., Hulme.W., Iliiev.I., Johnson.R.,
    Jones.C., Kamat.A., Kartas.A., Kells.C., LaRocque.K.,
    Lamazares.R., Lander.E., Lehoczyk.J., Levine.R., Liu.G.,
    MacLean.C., Macdonald.P., Major.J., Marquis.N., Matthews.C.,
    McCarthy.M., McEwan.P., McKernan.K., McPheeters.R., Meldrum.J.,
    Meneus.L., Mihova.T., Mlenga.V., Murphy.T., Naylor.J., Nguyen.C.,
    Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neill.D.,
    Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
    Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
    Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupbach.R.,
    Seaman.S., Severy.P., Sougniez.C., Spencer.B., Stange-Thomann.N.,
    Stojanovic.N., Strauss.N., Subramanian.A., Talamas.J., Testaye.S.,
    Theodore.J., Topham.K., Travers.M., Travis.N., Triggillo.J.,
    Vassiliou.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
    Young.G., Zainoun.J., Zembek.L., Zimmer.A. and Zody.M.
  Direct Submission
  Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L12706
  Center clone name: 679_G_21
  ----- Summary Statistics
  Sequencing vector: Plasmid; n/a; 100% of reads
  Chemistry: Dye-terminator Big Dye; version 0.960731
  Assembly program: Phrap; version 0.960731
  Consensus quality: 166430 bases at least Q40
  Consensus quality: 166426 bases at least Q30
  Consensus quality: 169078 bases at least Q20
  Insert size: 151000; agarose-fp
  Insert size: 169740; sum-of-contigs
  Quality coverage: 9.3 in Q20 bases; agarose-fp
  Quality coverage: 8.3 in Q20 bases; sum-of-contigs
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 9 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1
  * 744 843: gap of 100 bp
  * 844 1842: contig of 999 bp in length
  * 1843 1942: gap of 100 bp
  * 1943 4117: contig of 2175 bp in length
  * 4118 4217: gap of 100 bp
  * 4218 9901: contig of 5684 bp in length
  * 9902 10001: gap of 100 bp
  * 10002 22030: contig of 12029 bp in length
  * 22031 22130: gap of 100 bp
  * 22131 77732: contig of 55602 bp in length
  * 77733 77832: gap of 100 bp
  * 77833 110109: contig of 32277 bp in length
  * 110110 110209: gap of 100 bp
  * 110210 150387: contig of 40178 bp in length
  * 150388 150487: gap of 100 bp
  * 150488 170540: contig of 20053 bp in length.
  FEATURES
    Location/Qualifiers
      1..170540
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="11"
        /map="11"
        /clone="RP11-679G21"
        /clone_lib="RPC1-11 Human Male BAC"
      1..743
        /note="assembly fragment"
        clone_end:SP6
        vector_side:left"
    misc_feature
      844..1842

```

Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neill.D.,
 Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
 Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
 Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupbach.R.,
 Seaman.S., Severy.P., Sougniez.C., Spencer.B., Stange-Thomann.N.,
 Stojanovic.N., Strauss.N., Subramanian.A., Talamas.J., Testaye.S.,
 Theodore.J., Topham.K., Travers.M., Travis.N., Triggillo.J.,
 Vassiliou.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
 Young.G., Zainoun.J., Zembek.L., Zimmer.A. and Zody.M.
 Direct Submission
 Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L12706
 Center clone name: 679_G_21
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; version 0.960731
 Assembly program: Phrap; version 0.960731
 Consensus quality: 166430 bases at least Q40
 Consensus quality: 166426 bases at least Q30
 Consensus quality: 169078 bases at least Q20
 Insert size: 151000; agarose-fp
 Insert size: 169740; sum-of-contigs
 Quality coverage: 9.3 in Q20 bases; agarose-fp
 Quality coverage: 8.3 in Q20 bases; sum-of-contigs

TITLE
 JOURNAL
 COMMENT


```
/note="assembly_fragment"
1943. 4117
/note="assembly_fragment"
4218. 9901
/note="assembly_fragment"
10002. 22030
/note="assembly_fragment"
22131. 77732
/note="assembly_fragment"
77833. 110109
/note="assembly_fragment"
110210. 150387
/note="assembly_fragment"
150488. 170540
/note="assembly_fragment
clone_end:T7
vector_side:right"
SE COUNT 47282 a 39323 c 37471 g 45658 t 806 others
ORIGIN

Query Match 55.8%; Score 868; DB 2; Length 170540;
Best Local Similarity 100.0%; Pred. No. 1.3e-260;
Matches 868; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 687 AGGAGATGTGATCCGAGAAATTTAACTTAATGAGCTATACAGGGGCCCAAGAACTAT 746
Db 63433 AGGAGATGTGATCCGAGAAATTTAACTTAATGAGCTATACAGGGGCCCAAGAACTAT 63492

QY 747 CAAAGGCTGGAGACAATATCCCTGAGGAGCGCTGTGGCTTCAACGCCACCACAGTGT 806
Db 63493 CAAAGGCTGGAGACAATATCCCTGAGGAGCGCTGTGGCTTCAACGCCACCACAGTGT 63552

QY 807 CAGATGGGGAACAAGAAGGATAAATAAGATCTCTCACTTTGGCAGTGCTTCTCTCTG 866
Db 63553 CAGATGGGGAACAAGAAGGATAAATAAGATCTCTCACTTTGGCAGTGCTTCTCTCTG 63612

QY 867 TCAATTCAGGCTCTTTCCATACCAAGCGCTGAGCTGAGCTTCTTATTTATGTTT 926
Db 63613 TCAATTCAGGCTCTTTCCATACCAAGCGCTGAGCTGAGCTTCTTATTTATGTTT 63672

QY 927 CCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCTTCTGATTTTAAAGAGCGCATCTAGG 986
Db 63673 CCCTTTGGCTGTGACTGGGTGGGCGAGCATGCAGCTTCTGATTTTAAAGAGCGCATCTAGG 63732

QY 987 GAATTTGTCAGCCACCTACAGGAAGGCTGCCATGCTGTGGCCAACTGTTTCACTGGAGC 1046
Db 63733 GAATTTGTCAGCCACCTACAGGAAGGCTGCCATGCTGTGGCCAACTGTTTCACTGGAGC 63792

QY 1047 AAGAAAGAGATCTCATAGGACGGGGGAATGTTTCCCTCCAAGCTTGGGTCAGNGT 1106
Db 63793 AAGAAAGAGATCTCATAGGACGGGGGAATGTTTCCCTCCAAGCTTGGGTCAGNGT 63852

QY 1107 GTTAACCTCTTATCAGCTATTACAGATCTCCATGGTGTTCCTCCATGAACCTCTGTGGTT 1166
Db 63853 GTTAACCTCTTATCAGCTATTACAGATCTCCATGGTGTTCCTCCATGAACCTCTGTGGTT 63912

QY 1167 CATCATCTCTTATGTTGACCTGCACAGCTGTGTTAGACCTAGATTTAACCCCTAAGGTA 1226
Db 63913 CATCATCTCTTATGTTGACCTGCACAGCTGTGTTAGACCTAGATTTAACCCCTAAGGTA 63972

QY 1227 AGATGCTGGGTATAGAACGCTAGAAATTTTCCCCCAAGGACTCTGCTTCCCTTAAGGCC 1286
Db 63973 AGATGCTGGGTATAGAACGCTAGAAATTTTCCCCCAAGGACTCTGCTTCCCTTAAGGCC 64032

QY 1287 TTCTGGCTTCTGTTATGCTTCTTAAAGCTAATGAAGCTAAGCTTGTGCGTAGTCTTAA 1346
Db 64033 TTCTGGCTTCTGTTATGCTTCTTAAAGCTAATGAAGCTAAGCTTGTGCGTAGTCTTAA 64092

QY 1347 GGAGAAACCTTTAACCAACAAGTTTATCATTTGAAGACATATTTGAACACCCCTATT 1406
Db 64093 GGAGAAACCTTTAACCAACAAGTTTATCATTTGAAGACATATTTGAACACCCCTATT 64152

QY 1407 TTGTGGGATTTAGAACGGGTGAATAGAGGCTTTCCTTGTGTTGGTAGGACTT 1466

Db 64153 TTGTGGGATTTAGAACGGGTGAATAGAGGCTTTCCTTGTGTTGGTAGGACTT 64212
QY 1467 GGAGAGAAATCCCTCGACTTTCCTAACCTCTGACATACCTCCACACACAGTTGAT 1526
Db 64213 GGAGAGAAATCCCTCGACTTTCCTAACCTCTGACATACCTCCACACACAGTTGAT 64272
QY 1527 GGCCTTCGTAATAAAAAAGATTGGGATT 1554
Db 64273 GGCCTTCGTAATAAAAAAGATTGGGATT 64300

RESULT 13
AC021522/c
LOCUS AC021522 199992 bp DNA linear HTG 21-JUL-2001
DEFINITION Homo sapiens clone RP11-335H5, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC021522
VERSION AC021522.5 GI:14971311
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE
1 (bases 1 to 199992)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
2 (bases 1 to 199992)
Unpublished
Homo sapiens, clone RP11-335H5
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepey,Y., Collangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Garayna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 20, 2001 this sequence version replaced gi:14269795.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5504
Center clone name: 335_H_5
----- Summary Statistics
Sequencing vector: M13; M7815; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 192697 bases at least Q40
Consensus quality: 196336 bases at least Q30
Consensus quality: 197632 bases at least Q20
Insert size: 193000; agarose-fp
Insert size: 198692; sum-of-contents
Quality coverage: 8.3 in Q20 bases; agarose-fp
Quality coverage: 8.1 in Q20 bases.
```


* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 17413: contig of 17413 bp in length
* 17414 17513: gap of 100 bp
* 17514 18244: contig of 731 bp in length
* 18245 18344: gap of 100 bp
* 18345 19084: contig of 740 bp in length
* 19085 19184: gap of 100 bp
* 19185 20416: contig of 1232 bp in length
* 20417 20516: gap of 100 bp
* 20517 21625: contig of 1109 bp in length
* 21626 21725: gap of 100 bp
* 21726 102938: contig of 81213 bp in length
* 102939 103038: gap of 100 bp
* 103039 104905: contig of 1867 bp in length
* 104906 105005: gap of 100 bp
* 105006 108216: contig of 3211 bp in length
* 108217 108316: gap of 100 bp
* 108317 112875: contig of 4559 bp in length
* 112876 112975: gap of 100 bp
* 112976 122676: contig of 9701 bp in length
* 122677 122776: gap of 100 bp
* 122777 138821: contig of 16045 bp in length
* 138822 138921: gap of 100 bp
* 138922 168907: contig of 29986 bp in length
* 168908 169007: gap of 100 bp
* 169008 196861: contig of 27854 bp in length
* 196862 196961: gap of 100 bp
* 196962 199992: contig of 3031 bp in length.
  
```

FEATURES source

```

1. 199992
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="RPC1-11 Human Male BAC"
  1. 17413
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left
  misc_feature
    17514..18244
    /note="assembly_fragment"
  misc_feature
    18345..19084
    /note="assembly_fragment"
  misc_feature
    19185..20416
    /note="assembly_fragment"
  misc_feature
    20517..21625
    /note="assembly_fragment"
  misc_feature
    21726..102938
    /note="assembly_fragment"
  misc_feature
    103039..104905
    /note="assembly_fragment"
  misc_feature
    105006..108216
    /note="assembly_fragment"
  misc_feature
    108317..112875
    /note="assembly_fragment"
  misc_feature
    112976..122676
    /note="assembly_fragment"
  misc_feature
    122777..138821
    /note="assembly_fragment"
  misc_feature
    138922..168907
    /note="assembly_fragment"
  misc_feature
    169008..196861
    /note="assembly_fragment"
  misc_feature
    196862..199992
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right
  
```

BASE COUNT	54206 a	46381 c	45637 g	52463 t	1305 others
ORIGIN					
Query Match	55.8%	Score 868	DB 2	Length 199992	
Best Local Similarity	100.0%	Pred. No. 1.3e-260			
Matches 868	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
Qy	687	AGGAGATGTGATCCGAGAAATTAACATTAATGAGCTATACACGCGGCCCAAGAACTAT	746		
Db	183496	AGGAGATGTGATCCGAGAAATTAACATTAATGAGCTATACACGCGGCCCAAGAACTAT	183437		
Qy	747	CAAGAGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACCCCAACACAGTGT	806		
Db	183436	CAAGAGCTGGAGACAATATCCCTGAGGAGCAGCCTGTGGCTTCAACCCCAACACAGTGT	183377		
Qy	807	CAGATGGGAAAACAAGAGGATAAATAAGATCCTCACTTTGGCAGTGCTTCTCTCTCTG	866		
Db	183376	CAGATGGGAAAACAAGAGGATAAATAAGATCCTCACTTTGGCAGTGCTTCTCTCTCTG	183317		
Qy	867	TCAATTCAGGCTCTTTCCATAACCAAGCCTCAGGCTGCAGCCTTTATTTATGTTTT	926		
Db	183316	TCAATTCAGGCTCTTTCCATAACCAAGCCTCAGGCTGCAGCCTTTATTTATGTTTT	183257		
Qy	927	CCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCACTTCTTGATTTAAAGAGGCATCTAGG	986		
Db	183256	CCCTTTGGCTGTGACTGGGTGGGCGCAGCATGCACTTCTTGATTTAAAGAGGCATCTAGG	183197		
Qy	987	GAATTGTCAAGCCCTACAGGAAGGCTGCCATGCTGTGGCCAACTGTTTCACCTGGAGC	1046		
Db	183196	GAATTGTCAAGCCCTACAGGAAGGCTGCCATGCTGTGGCCAACTGTTTCACCTGGAGC	183137		
Qy	1047	AAGAAAGAGATCTCATAGGAGCGGGGAAATGGTTTCCCTCCAAGCTTGGGTCAAGTGT	1106		
Db	183136	AAGAAAGAGATCTCATAGGAGCGGGGAAATGGTTTCCCTCCAAGCTTGGGTCAAGTGT	183077		
Qy	1107	GTTAACTGCTTATCAGCTATTTCAGACATCTCCATGGTTTCTCCATGAACCTCTCTGGTTT	1166		
Db	183076	GTTAACTGCTTATCAGCTATTTCAGACATCTCCATGGTTTCTCCATGAACCTCTCTGGTTT	183017		
Qy	1167	CATCATTCCTTCTTAGTTAGCTGCACAGCTGGTTAGACCTAGATTTAACCCCTAAGGTA	1226		
Db	183016	CATCATTCCTTCTTAGTTAGCTGCACAGCTGGTTAGACCTAGATTTAACCCCTAAGGTA	182957		
Qy	1227	AGATGCTGGGTATAGAACGCTAAGAAATTTCCCCCAAGGACTCTCTCTCTTAAGCCC	1286		
Db	182956	AGATGCTGGGTATAGAACGCTAAGAAATTTCCCCCAAGGACTCTCTCTCTTAAGCCC	182897		
Qy	1287	TTCTGGCTTCTTATGGTCTTTCATTAAGATATAAGCCTAACTTTTCTCGCTAGTCTAA	1346		
Db	182896	TTCTGGCTTCTTATGGTCTTTCATTAAGATATAAGCCTAACTTTTCTCGCTAGTCTAA	182837		
Qy	1347	GGAGAAACCTTTAACCAACAAGTTTTTATCATTAAGACAAATATTTGAACACCCCTATT	1406		
Db	182836	GGAGAAACCTTTAACCAACAAGTTTTTATCATTAAGACAAATATTTGAACACCCCTATT	182777		
Qy	1407	TTGTGGGATTGAGAGGGGTGATAGAGGCTTGAGACTTTCCTTTCTGTGGTAGGACTT	1466		
Db	182776	TTGTGGGATTGAGAGGGGTGATAGAGGCTTGAGACTTTCCTTTCTGTGGTAGGACTT	182717		
Qy	1467	GGAGGAGAAATCCCTCGACTTTTCACTAACCTCTGACATCTCCACACACCCAGTGTAT	1526		
Db	182716	GGAGGAGAAATCCCTCGACTTTTCACTAACCTCTGACATCTCCACACACCCAGTGTAT	182657		
Qy	1527	GGCTTTCCGTAATAAAGATTTGGGATT	1554		
Db	182656	GGCTTTCCGTAATAAAGATTTGGGATT	182629		

RESULT 14
 AP000727/c
 LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-679G21 map 11q12, WORKING
 DRAFT SEQUENCE, 40 unordered pieces.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

AP000727
 AP000727.2 GI-8118896
 HTG; HTGS-PHASE1; HTGS-DRAFT.
 Homo sapiens DNA, clone:RP11-679G21.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 153394)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 153,394 genomic DNA of 11q12
 Published Only in DataBase (1999)
 2 (bases 1 to 153394)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (16-NOV-1999) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gsc.riken.go.jp)
 URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 On May 31, 2000 this sequence version replaced gi:6997582.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: RP11-679G21
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 133607 bases at least Q40
 Consensus quality: 142217 bases at least Q30
 Consensus quality: 146720 bases at least Q20
 Insert size: 149494; sum-of-contigs
 Quality coverage: 4.03x in Q20 bases; sum-of-contigs

COMMENT

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 40 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 11871: contig of 11871 bp in length
 * 11872 11971: gap of 100 bp
 * 11972 22347: contig of 10376 bp in length
 * 22348 22447: gap of 100 bp
 * 22448 33540: contig of 11093 bp in length
 * 33541 33640: gap of 100 bp
 * 33641 41577: contig of 7937 bp in length
 * 41578 41677: gap of 100 bp
 * 41678 45975: contig of 4298 bp in length
 * 45976 46075: gap of 100 bp
 * 46076 53127: contig of 7052 bp in length
 * 53128 53227: gap of 100 bp
 * 53228 60314: contig of 7087 bp in length
 * 60315 60414: gap of 100 bp
 * 60415 65544: contig of 5130 bp in length
 * 65545 65644: gap of 100 bp
 * 65645 71027: contig of 5383 bp in length
 * 71028 71127: gap of 100 bp
 * 71128 75422: contig of 4295 bp in length
 * 75423 75522: gap of 100 bp
 * 75523 79274: contig of 3752 bp in length
 * 79275 79374: gap of 100 bp
 * 79375 84398: contig of 5024 bp in length
 * 84399 84498: gap of 100 bp
 * 84499 88990: contig of 4492 bp in length
 * 88991 89090: gap of 100 bp
 * 89091 93436: contig of 4346 bp in length
 * 93437 93536: gap of 100 bp
 * 93537 96655: contig of 3119 bp in length
 * 96656 96755: gap of 100 bp
 * 96756 99175: contig of 2420 bp in length
 * 99176 99275: gap of 100 bp
 * 99276 102659: contig of 3384 bp in length
 * 102660 102759: gap of 100 bp
 * 102760 105824: contig of 3065 bp in length
 * 105825 105924: gap of 100 bp
 * 105925 109361: contig of 3437 bp in length
 * 109362 109461: gap of 100 bp
 * 109462 112303: contig of 2842 bp in length
 * 112304 112403: gap of 100 bp
 * 112404 115484: contig of 3081 bp in length
 * 115485 115584: gap of 100 bp
 * 115585 118972: contig of 3388 bp in length
 * 118973 119072: gap of 100 bp
 * 119073 121620: contig of 2548 bp in length
 * 121621 121720: gap of 100 bp
 * 121721 124322: contig of 2602 bp in length
 * 124323 124422: gap of 100 bp
 * 124423 127088: contig of 2666 bp in length

NOTE: This is a 'working draft' sequence. It currently consists of
 40 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

1 11871 contig of 11871 bp in length
 11972 22347 contig of 10376 bp in length
 22448 33540 contig of 11093 bp in length
 33641 41577 contig of 7937 bp in length
 41678 45975 contig of 4298 bp in length
 46076 53127 contig of 7052 bp in length
 53228 60314 contig of 7087 bp in length
 60415 65544 contig of 5130 bp in length
 65645 71027 contig of 5383 bp in length
 71128 75422 contig of 4295 bp in length
 75523 79274 contig of 3752 bp in length
 79375 84398 contig of 5024 bp in length
 84499 88990 contig of 4492 bp in length
 89091 93436 contig of 4346 bp in length
 93537 96655 contig of 3119 bp in length
 96756 99175 contig of 2420 bp in length
 99276 102659 contig of 3384 bp in length
 102760 105824 contig of 3065 bp in length
 105925 109361 contig of 3437 bp in length
 109462 112303 contig of 2842 bp in length
 112404 115484 contig of 3081 bp in length
 115585 118972 contig of 3388 bp in length
 119073 121620 contig of 2548 bp in length
 121721 124322 contig of 2602 bp in length
 124423 127088 contig of 2666 bp in length

```

* 127089 127188: gap of 100 bp
* 127189 130306: contig of 3118 bp in length
* 130307 130406: gap of 100 bp
* 130407 131809: contig of 1403 bp in length
* 131810 131909: gap of 100 bp
* 131910 133740: contig of 1831 bp in length
* 133741 133840: gap of 100 bp
* 133841 136109: contig of 2269 bp in length
* 136110 136209: gap of 100 bp
* 136210 137539: contig of 1330 bp in length
* 137540 137639: gap of 100 bp
* 137640 139390: contig of 1751 bp in length
* 139391 139490: gap of 100 bp
* 139491 140939: contig of 1449 bp in length
* 140940 141039: gap of 100 bp
* 141040 142493: contig of 1454 bp in length
* 142494 142593: gap of 100 bp
* 142594 144234: contig of 1641 bp in length
* 144235 144334: gap of 100 bp
* 144335 145850: contig of 1516 bp in length
* 145851 145950: gap of 100 bp
* 145951 147405: contig of 1455 bp in length
* 147406 147505: gap of 100 bp
* 147506 149002: contig of 1497 bp in length
* 149003 149102: gap of 100 bp
* 149103 150523: contig of 1421 bp in length
* 150524 150623: gap of 100 bp
* 150624 152235: contig of 1612 bp in length
* 152236 152335: gap of 100 bp
* 152336 153394: contig of 1059 bp in length.
FEATURES
    Location/Qualifiers
        source
            1..153394
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="11"
                /map="11q12"
                /clone="RP11-679621"
            1..11871
                /note="assembly_fragment"
            11972..22347
                /note="assembly_fragment"
            22448..33540
                /note="assembly_fragment"
            33641..41577
                /note="assembly_fragment"
            41678..45975
                /note="assembly_fragment"
            46076..53127
                /note="assembly_fragment"
            53228..60314
                /note="assembly_fragment"

```

ery Match 55.0%; Score 856; DB 2: Length 153394;

est Local Similarity 99.9%; Pred. No. 7.4e-257;

Matches 867; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 687 AGAGAATGTGATCCGAGATTTAACTTAATGAGCTATACACGGCGGCAAGAACTAT 746
|||||
DB 11478 AGAGAATGTGATCCGAGATTTAACTTAATGAGCTATACACGGCGG-CAAGAACTAT 11420
|||||
QY 747 CAAAGCTCGAGACAATATCCCTGAGGAGCGCTGTGGCTTCAACCCCAACACAGCTGT 806
|||||
DB 11419 CAAAGCTCGAGACAATATCCCTGAGGAGCGCTGTGGCTTCAACCCCAACACAGCTGT 11360
|||||
QY 807 CAGATGGGAAACAAAGAGGATAAATAGATCCTCAGCTTTGGCAGTCTCTCTCCTG 866
|||||
DB 11359 CAGATGGGAAACAAAGAGGATAAATAGATCCTCAGCTTTGGCAGTCTCTCTCCTG 11300
|||||
QY 867 TCAATTCAGGCTCTTTCCATACACCAAGCCCTGAGGCTGCACGCTTTTATATGTTTT 926
|||||
DB 11299 TCAATTCAGGCTCTTTCCATACACCAAGCCCTGAGGCTGCACGCTTTTATATGTTTT 11240
|||||
QY 927 CCTTTGGCTGTGACTGTGGTGGGCGAGCATGCGAGCTCTGATTTTAAAGAGGCATCTAGG 986

```

```

DB 11239 CCCTTTGGCTGTGACTGTGGGTGGGCGAGCATGCGAGCTCTGATTTTAAAGAGCATCTAGG 11180
|||||
QY 987 GAATTTGTAGGCAACCTTACAGGAAGGCTGCCATGCTGTGGCAACTGTTTCACTGAGGC 1046
|||||
DB 11179 GAATTTGTAGGCAACCTTACAGGAAGGCTGCCATGCTGTGGCAACTGTTTCACTGAGGC 11120
|||||
QY 1047 AAGAAAGAGATCTCATAGGAGCGGGGAAATGGTTTCCCTCCAAGCTTGGGTCACTGT 1106
|||||
DB 11119 AAGAAAGAGATCTCATAGGAGCGGGGAAATGGTTTCCCTCCAAGCTTGGGTCACTGT 11060
|||||
QY 1107 GTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTCCATGAACCTCTCTGTTT 1166
|||||
DB 11059 GTTAACTGCTTATCAGCTATTTCAGACATCTCCATGTTTCTCCATGAACCTCTCTGTTT 11000
|||||
QY 1167 CATCATTCCTTCTTAGTTGACCTGCGACAGCTTGGTTAGACCTAGATTAAACCTTAAGTA 1226
|||||
DB 10999 CATCATTCCTTCTTAGTTGACCTGCGACAGCTTGGTTAGACCTAGATTAAACCTTAAGTA 10940
|||||
QY 1227 AGATGCTGGGTATAGAACGCTTAAGAAATTTTCCCCCAAGGACTCTTCTCTTCCCTTAAGCCC 1286
|||||
DB 10939 AGATGCTGGGTATAGAACGCTTAAGAAATTTTCCCCCAAGGACTCTTCTCTTCCCTTAAGCCC 10880
|||||
QY 1287 TTCTGGCTTCGTTTATGTTTATGTTTATGTTTAAAGTATTAAGCCTTAAGCTTCTCGCTAGCTAA 1346
|||||
DB 10879 TTCTGGCTTCGTTTATGTTTATGTTTATGTTTAAAGTATTAAGCCTTAAGCTTCTCGCTAGCTAA 10820
|||||
QY 1347 GGAGAAACCTTTAACCAAAAGTTTATGTTTATGTTTAAAGTATTAAGCCTTAAGCTTCTCGCTAGCTAA 1406
|||||
DB 10819 GGAGAAACCTTTAACCAAAAGTTTATGTTTATGTTTAAAGTATTAAGCCTTAAGCTTCTCGCTAGCTAA 10760
|||||
QY 1407 TTCTGGGATTTAGAGAGGGGTGATAGAGGCTTGTAGACTTTCCTTTCTGTTGGTAGGACTT 1466
|||||
DB 10759 TTCTGGGATTTAGAGAGGGGTGATAGAGGCTTGTAGACTTTCCTTTCTGTTGGTAGGACTT 10700
|||||
QY 1467 GGAGGAGAAATCCCTGGACTTTCACATAACCTCTGACATACCTCCCAACACCAAGCTTGTAT 1526
|||||
DB 10699 GGAGGAGAAATCCCTGGACTTTCACATAACCTCTGACATACCTCCCAACACCAAGCTTGTAT 10640
|||||
QY 1527 GGCTTTCCGTAATAAAAGATTGGGATT 1554
|||||
DB 10639 GGCTTTCCGTAATAAAAGATTGGGATT 10612
|||||
RESULT 15
AC013381
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-353M23 map 11, LOW-PASS
SEQUENCE SAMPLING.
AC013381
VERSION AC013381.4 GI:9123907
KEYWORDS HTG; HTGS_PHASED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166341)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-353M23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166341)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collins,S., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Tessfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 166341)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Codymore, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galaan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, B., Olivari, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910757.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3769
Center clone name: 353_M_23
-----
* NOTE: This record contains 174 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 845 944: gap of 100 bp in length
* 945 1799: contig of 855 bp in length
* 1800 1899: gap of 100 bp
* 1900 2775: contig of 876 bp in length
* 2776 2875: gap of 100 bp
* 2876 3771: contig of 896 bp in length
* 3772 3871: gap of 100 bp
* 3872 4761: contig of 890 bp in length
* 4762 4861: gap of 100 bp
* 4862 5736: contig of 875 bp in length
* 5737 5836: gap of 100 bp
* 5837 6640: contig of 804 bp in length
* 6641 6740: gap of 100 bp
* 6741 7617: contig of 877 bp in length
* 7618 7717: gap of 100 bp
* 7718 8596: contig of 879 bp in length
* 8597 8696: gap of 100 bp
* 8697 9555: contig of 859 bp in length
*
* 9556 9655: gap of 100 bp
* 9656 10517: contig of 862 bp in length
* 10518 10617: gap of 100 bp
* 10618 11478: contig of 861 bp in length
* 11479 11578: gap of 100 bp
* 11579 12420: contig of 842 bp in length
* 12421 12520: gap of 100 bp
* 12521 13422: contig of 902 bp in length
* 13423 13522: gap of 100 bp
* 13523 14382: contig of 860 bp in length
* 14383 14482: gap of 100 bp
* 14483 15347: contig of 865 bp in length
* 15348 15447: gap of 100 bp
* 15448 16303: contig of 856 bp in length
* 16304 16403: gap of 100 bp
* 16404 17269: contig of 866 bp in length
* 17270 17369: gap of 100 bp
* 17370 18239: contig of 870 bp in length
* 18240 18339: gap of 100 bp
* 18340 19192: contig of 853 bp in length
* 19193 19292: gap of 100 bp
* 19293 20160: contig of 868 bp in length
* 20161 20260: gap of 100 bp
* 20261 21120: contig of 860 bp in length
* 21121 21220: gap of 100 bp
* 21221 22076: contig of 856 bp in length
* 22077 22176: gap of 100 bp
* 22177 23040: contig of 864 bp in length
* 23041 23140: gap of 100 bp
* 23141 24000: contig of 860 bp in length
* 24001 24100: gap of 100 bp
* 24101 24966: contig of 866 bp in length
* 24967 25066: gap of 100 bp
* 25067 25954: contig of 888 bp in length
* 25955 26054: gap of 100 bp
* 26055 26916: contig of 862 bp in length
* 26917 27016: gap of 100 bp
* 27017 27891: contig of 875 bp in length
* 27892 27991: gap of 100 bp
* 27992 28859: contig of 868 bp in length
* 28860 28959: gap of 100 bp
* 28960 29827: contig of 868 bp in length
* 29828 29927: gap of 100 bp
* 29928 30828: contig of 901 bp in length
* 30829 30928: gap of 100 bp
* 30929 31742: contig of 814 bp in length
* 31743 31842: gap of 100 bp
* 31843 32699: contig of 857 bp in length
* 32700 32799: gap of 100 bp
* 32800 33662: contig of 863 bp in length
* 33663 33762: gap of 100 bp
* 33763 34638: contig of 876 bp in length
* 34639 34738: gap of 100 bp
* 34739 35603: contig of 865 bp in length
* 35604 35703: gap of 100 bp
* 35704 36578: contig of 875 bp in length
* 36579 36678: gap of 100 bp
* 36679 37519: contig of 841 bp in length
* 37520 37619: gap of 100 bp
* 37620 38483: contig of 864 bp in length
* 38484 38583: gap of 100 bp
* 38584 39427: contig of 844 bp in length
* 39428 39527: gap of 100 bp
* 39528 40387: contig of 860 bp in length
* 40388 40487: gap of 100 bp
* 40488 41342: contig of 855 bp in length
* 41343 41442: gap of 100 bp
* 41443 42291: contig of 849 bp in length
* 42292 42391: gap of 100 bp
* 42392 43246: contig of 855 bp in length
* 43247 43346: gap of 100 bp
* 43347 44206: contig of 860 bp in length
* 44207 44306: gap of 100 bp

```

TITLE

COMMENT

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 11:52:55 ; Search time 40 Seconds
(without alignments)
620.067 Million cell updates/sec

Title: US-09-954-846-2

Perfect score: 1359

Sequence: 1 MAVLAPLIALVSPRLSRW.....EQPVASTPTTVSDGENKKDK 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR.73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	391	28.8	265	2	S72570
2	173	12.7	253	2	T48523
3	125	9.2	135	2	C90393
4	124	9.1	104	1	TXBY1
5	122.5	9.0	117	2	E70107
6	122	9.0	103	1	TXBY2
7	121.5	8.9	145	2	AD0397
8	121	8.9	664	2	S44756
9	120.5	8.9	105	2	H69517
10	120.5	8.9	134	2	C69410
11	119	8.8	491	2	S71862
12	119	8.8	493	2	T34092
13	117.5	8.6	115	2	T29044
14	116	8.5	250	2	E87921
15	116	8.5	606	2	T31557
16	115.5	8.5	359	2	T30644
17	114	8.4	102	2	S73896
18	114	8.4	139	2	G70464
19	112.5	8.3	359	2	S73880
20	112	8.2	105	2	AG2042
21	111.5	8.2	127	2	B91218
22	111.5	8.2	127	2	C86064
23	111	8.2	643	1	S32476
24	110.5	8.1	104	2	S77780
25	110	8.1	108	1	S02802
26	110	8.1	122	2	T04090
27	110	8.1	364	2	T09614
28	110	8.1	410	2	T25574
29	109	8.0	584	2	S06318

30 109 8.0 638 1 ISMSER protein disulfide-
31 108.5 8.0 105 2 S76386 thioedoxin M-1 -
32 108.5 8.0 139 2 E65036 rotable thioedoxi
33 108.5 8.0 139 2 G85904 probable thioedox
34 108.5 8.0 139 2 H91059 probable thioedox
35 108.5 8.0 645 1 A23723 protein disulfide-
36 107.5 7.9 144 2 B82423 thioedoxin 2 VCAO
37 107 7.9 1023 2 G96509 protein F27F5.21
38 106.5 7.8 443 2 D86183 hypothetical prote
39 106.5 7.8 496 2 A54757 protein disulfide-
40 106 7.8 102 1 G64213 thioedoxin - Myco
41 106 7.8 102 2 D72052 thioedoxin CP0088
42 106 7.8 102 2 H86572 thioedoxin [impor
43 106 7.8 116 2 B70851 probable trxC prot
44 105.5 7.8 126 1 S18590 thioedoxin h1 - c
45 105.5 7.8 788 2 T26967 hypothetical prote

ALIGNMENTS

RESULT 1

S72570

hypothetical protein C35D10.1n - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 31-Oct-1997

C.Accession: S72570

R.Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
ulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnston, L.;
B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.
submitted to the EMBL Data Library, February 1995
A.Authors: Shownkeen, R.; Smaldon, N.; Smith, A.; Sonhammer, E.; Staden, R.; Sulston
proat, J.; Wohlman, P.

A.Description: The C. elegans genome project: Contiguous nucleotide sequence of over
A.Reference number: S72566

A.Accession: S72570

A.Molecule type: DNA

A.Residues: 1-265 <WIL>

A.Cross-references: EMBL:U21324; NID:G687879; PID:G687889

A.Experimental source: strain Bristol N2

C.Genetics:

A.Map position: 3

A.Introns: 52/3; 111/1; 213/2

A.Note: C35D10.10

Query Match 28.8%; Score 391; DB 2; Length 265;

Best Local Similarity 32.2%; Pred. No. 5.4e-28;

Matches 84; Conservative 52; Mismatches 81; Indels 44; Gaps 6;

QY 14 VPRLS--RWLAQPYLLSALLSAFLVRLKLPPLCHGLPTOREDNPCDFDHWREVEILMF 71

Db 3 IPRLDVRRALTAFHFFNTLLALAPPVIRS-TSLGDYV-FAVEGNEQCEIDSREREILMF 60

QY 72 LSAIVMMKNRS-----MELMTC-----KPL 93

Db 61 LLILAWKGRATNMVHVNIFLFSKIAGMFLTRADILPGIITLACILVTLVLPPEV 120

QY 94 YMGPEYIKYFNKDTIDEELERDKRYTWIVFFANWNCQSFAPYADLSKYNCTGNLF 153

Db 121 YNGPEQVTFQGEQLFEELTRNRTIWIQFTTWSPECRHTSPVFAELSKQFTLPNKKF 180

QY 154 GKVDVGRYTVSTRYKYSTSLTKQLPTLILFQGGKEAMRRPQIDKKGRAVSWTSEENV 213

Db 181 KGLDIGRWAKEGERFVRNHPMSRQLPTICVFKDAKEIARRPLVNDSSRAVPFVFESENC 240

QY 214 IREFNLNELY--QRAKKLSKA 232

Db 241 VLAFDLLNLYNEQKEKKGAKA 261

RESULT 2

T48523

hypothetical protein T22p22.30 - Arabidopsis thaliana

DB 87 TEESQDIANRYGIMS-----LPTINFFKNGE 112

RESULT 4
TXBY1

N:thioredoxin I - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G7746; protein YGR209c; thioredoxin 2
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jun-2000
C:C-Accession: S15049; B39847; S05793; S53932; S61947; S64531; S63858; A38669
R:Can, Z.R.
J. Biol. Chem. 266, 1692-1696, 1991
A:Title: Yeast thioredoxin genes.
A:Reference number: A38669; MUID:91107668; PMID:1988444
A:Accession: S15049
A:Molecule type: DNA
A:Residues: 1-104 <CAN>
A:Cross-references: GB:M59168; NID:gl73025; PIDN:AAA35170.1; PID:gl73026
R:Muller, E.G.D.
J. Biol. Chem. 266, 9194-9202, 1991
A:Title: Thioredoxin deficiency in yeast prolongs S phase and shortens the G1 interval.
A:Reference number: A39847; MUID:91225027; PMID:2026619
A:Accession: B39847
A:Molecule type: DNA
A:Residues: 1-104 <MUL>
A:Cross-references: GB:M62648; NID:gl73049; PIDN:AAA35178.1; PID:gl73050
R:Hall, D.E.; Baldesten, A.; Holmgren, A.; Reichard, P.
Eur. J. Biochem. 23, 328-335, 1971
A:Title: Yeast thioredoxin. Amino-acid sequence around the active-center disulfide of
A:Reference number: S05793; MUID:72100583; PMID:945270
A:Accession: S05793
A:Molecule type: Protein
A:Residues: 2:27-43; 98-104 <HAL>
A:Note: the sequence from the summary and from Fig. 5 is inconsistent with that from
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-P
submitted to the EMBL Data Library, April 1995
A:Description: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome V
es, of the human.
A:Reference number: S53922
A:Accession: S53932
A:Molecule type: DNA
A:Residues: 1-104 <GUE>
A:Cross-references: EMBL:249133; NID:g790489; PIDN:CAA89002.1; PID:g790500
A:Experimental source: strain S288C
R:Song, J.M.; Cheung, E.; Rabinowitz, J.C.
submitted to the EMBL Data Library, November 1995
A:Description: Analysis of the 15.6-kb fragment encompassing the ADE3 gene.
A:Reference number: S61947
A:Accession: S61947
A:Molecule type: DNA
A:Residues: 1-104 <SON>
A:Cross-references: EMBL:U40843; NID:gl165213; PIDN:AAA85584.1; PID:gl165214
A:Experimental source: strain GRF88
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-P
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64517
A:Accession: S64531
A:Molecule type: DNA
A:Residues: 1-104 <GUW>
A:Cross-references: EMBL:272994; NID:gl323374; PIDN:CAA97236.1; PID:gl323375; GSPDB:G
A:Experimental source: strain S288C
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-P
Yeast 12, 273-280, 1996
A:Title: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII rev
terial electron-transferring flavoproteins (beta-chain) and of the Escherichia coli p
A:Reference number: S63848; MUID:97060019; PMID:8904340
A:Accession: S63858
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-104 <GUF>
A:Cross-references: EMBL:249133; NID:g790489; PIDN:CAA89002.1; PID:g790500
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:

[illegible]

A:Residues: 1-103 <GAN>
A:Cross-references: EMBL:M59169; NID:g173027; PIDN:AAA35171.1; PID:g173028
R:Mueller, E.G.D.
J. Biol. Chem. 266, 9194-9202, 1991
A:Title: Thioredoxin deficiency in yeast prolongs S phase and shortens the G1 interval
A:Reference number: A39847; MUID:91225027; PMID:2026619
A:Accession: A39847
A:Molecule type: DNA
A:Residues: 1-103 <ML>
A:Cross-references: GB:M62647; NID:g173047; PIDN:AAA35177.1; PID:g173048
R:Hall, D.E.; Baldsten, A.; Holmgren, A.; Reichard, P.
Eur. J. Biochem. 23, 328-335, 1971
A:Title: Yeast thioredoxin. Amino-acid sequence around the active-center disulfide of
A:Reference number: S05793; MUID:72100583; PMID:4945270
A:Accession: S15360
A:Molecule type: protein
A:Residues: 25-34 <HAL>
R:Koetter, P.; Rose, M.; Entian, K.D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64863
A:Accession: S64870
A:Molecule type: DNA
A:Residues: 1-103 <KOE>
A:Cross-references: EMBL:Z73215; NID:g1360372; PIDN:CAA97572.1; PID:g1360373; GSPDB:S3
A:Note: experimental_source strain S288C
C:Genetics:
A:Gene: SGD:TRX1; TR-II; MIPS:YLR043C
A:Cross-references: SGD:S0004033; MIPS:YLR043C
A:Map position: 12R
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
P:2-103/Product: thioredoxin II #status predicted <MAT>
F:9-90/Domain: thioredoxin homology <THR>
F:30-33/Disulfide bonds: redox-active #status experimental

Query Match 9.0%; Score 122; DB 1; Length 103;
Best Local Similarity 34.1%; Pred. No. 0.0005;
Matches 29; Conservative 17; Mismatches 29; Indels 10; Gaps 3;

Qy 109 DEELERKRVTWIVFFANNSDCQSFAPYADLSKYNTGLNFQVVDGRTDYSTRY 168
I : ||| :||:| | : ||| : ||| : ||| : ||| :
Db 12 DSIAIQDKLV--VVDFYATWCGCKMIAPKEFSEQY--PQADFYLKDVLGDVAQKN 67
I : ||| :||:| | : ||| : ||| : ||| : ||| :
Qy 169 KVSTSPLTQKLPTLLIFQGGKEAMR 193
I : ||| :||:| | : ||| : ||| : ||| : ||| :
Db 68 EVSA-----MPTLLLFNGKEYAK 86
I : ||| :||:| | : ||| : ||| : ||| : ||| :

RESULT 7
AD0397
Thioredoxin 2 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AD0397
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,
I., M.; Rutherford, K.; Simmonds, T.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0397
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92504.1; PID:g15981204; GSPDB:GN00175
C:Genetics:
A:Gene: trxC
C:Superfamily: thioredoxin; thioredoxin homology

Query Match 8.9%; Score 121.5; DB 2; Length 145;
Best Local Similarity 27.7%; Pred. No. 0.00088;
Matches 33; Conservative 30; Mismatches 33; Indels 23; Gaps 6;

Qy 106 KTIDELERDKRVTWIVFFANWDCQSFAPYADLSLKYNTGLNFGKVDVGRYTDVS 165
 Db 43 ETDLKLLQDD--LPWVIDWAPWCGCFRSPFAPFAETAEE--RAGKVRFRVKVNTAEAPALS 99
 Qy 166 TRYKVVSTPLTKOLPTLLIFOGSGKEAMRRPQIDKKGRAV-----SWTFSEENVIREN 218
 Db 100 TRFRI-----RSIPTIMLYRNGK-----MIDMLNGAVPKAPEDNWL--DEQLSRDPN 144

RESULT 8
 S44756
 Probable protein disulfide-isomerase (EC 5.3.4.1) - Caenorhabditis elegans
 N:Alternate names: C14B9.2 protein
 C:Species: Caenorhabditis elegans
 C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 16-Jul-1999
 C:Accession: S44756
 R:Favello, A.D.
 submitted to the EMBL Data Library, May 1993
 A:Description: Sequence of the C. elegans cosmid C14B9.
 A:Reference number: S44617
 A:Accession: S44756
 A:Status: preliminary
 Molecule type: DNA
 Residues: 1-864 <FAV>
 ..:Cross-references: EMBL:L15188; NID:g289640; PIDN:AAA27952.1; PID:g289643
 C:Genetics:
 A: Introns: 43/2; 110/2; 631/3
 C:Superfamily: protein disulfide-isomerase; thioredoxin homology
 C:Keywords: intramolecular oxidoreductase; isomerase
 F:90-170/Domain: thioredoxin homology <TX1>
 F:201-285/Domain: thioredoxin homology <TX2>
 F:553-640/Domain: thioredoxin homology <TX3>

Query Match 8.9%; Score 121; DB 2; Length 664;
 Best Local Similarity 25.1%; Pred. No. 0.0073;
 Matches 51; Conservative 22; Mismatches 72; Indels 58; Gaps 9;

Qy 48 GLPTQR--EDG-NPCDF-----WREVEITLFLSAIVMMKNRRSMFLMTCKPLPY 94
 Db 151 GYPTLKEWKDGKPNYDGGDEAGIVEWVE-----SRVDPNYKPP-- 191
 Qy 95 MGPEYIKFNDKTYIDELERDKRVTWIVFFANWDCQSFAPYADLSLKYNTG--LN 152
 Db 192 --PEEVTLTTFNDFDIFNNELV--LVEFYAPWCGHCKLAPEYERAAQKLKAQGSVKV 247
 Qy 153 FGKVDVGRYTDVSTRYKVVSTPLTKOLPTLLIFOGSGKEAMRRPQIDKKGRAVSWTFSEEN 212
 Db 248 LGKVDATITEKDLGTYKGVSGYPTMK-----IIRNGRRFDYNGPREA 288
 Qy 213 V-IREFNLNELYQRAKKLSKAGD 234
 Db 289 AGIIKYMTDQSKPAKKLPKLKD 311

RESULT 9
 H69517
 Thioredoxin (trx-4) homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
 C:Accession: H69517
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 : Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.;
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69517
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-105 <KLE>
 A:Cross-references: GB:AE000956; GB:AE000782; NID:g2689279; PIDN:AAB89115.1; PID:g264838

	Query Match	8.6%	Score 117.5;	DB 2;	Length 115;
	Best Local Similarity	29.2%;	Pred. No. 0.0015;		
	Matches	26;	Conservative	20;	Mismatches 34; Indels 9; Gaps 3;
Qy	100 IKFVNKDTDEEELERKRVTWIEVFANNSQCQSFAPYIADLSLKYNCITGLNFGKVDVG	159			
	: : : :	:	:	:	:
Dd	11 VKYF-QSDFEQLRHQHPEKIILDFYATWCGPCAIAPLYKELATTHK--GIIFCKVDVD	67			
	: : : :	:	:	:	:
Qy	160 RYTNDVSTRYKVTSPLTKQLPTLILFQQG	188			
	: : : :	:	:	:	:
Dd	68 EAEEDLCSKKYDV-----KMMPTEIFIKNG	90			
	: : : :	:	:	:	:

Query Match	8.5%;	Score 116;	DB 2;	Length 250;
Best Local Similarity	25.3%;	Pred. No. 0.0057;		
Matches	37;	Conservative	21;	Mismatches 48; Indels 40; Gaps 7;
QY	112	LER-DKRVTVIIEFFANWNSDQSFAPIVADLSKY-NCTGLNFGKVDVGRVTDVSTRYK	169	
Db	16	LERSDANRLIIIDFFANWGPCRMISPIEQFSAEYGNAT---FLKVNCDVARDIVQRN	72	
QY	170	VSTSPLTQOLPILILFQGGKEAMRRPQIOKKGRANSWTFSEENVIREFNINELYQRAKKL	229	
Db	73	ISA-----MPTFIFLK-----NRQVD-----MVRGANQQAIAEKIRQ-	105	
QY	230	SKAGDNIPEEQPVASTPTTVSDGENK	255	
Db	106	-----HYSPTPANPNAASDEKR	123	

RESULT 15
T31557
hypothetical protein F5G4.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence revision 29-Oct-1999 #1

C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000

Search completed: July 9, 2003, 11:58:00
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 9, 2003, 11:39:55 ; Search time 11 seconds
(without alignments)
972.808 Million cell updates/sec

Title: US-09-954-846-2

Perfect score: 1359

Sequence: 1 MAVLAPLIALVSVPLSRW.....EQPVASTPTVSDGENKKDK 258

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	9.1	103	1	TRX1_YEAST
2	122.5	9.0	117	1	THIO_BORBU
3	122	9.0	102	1	TRX2_YEAST
4	121	8.9	618	1	PD44_CAEEL
5	117.5	8.6	115	1	TH11_CAEEL
6	116	8.5	127	1	TH10_NEUCR
7	114	8.4	102	1	THIO_MYCPM
8	113	8.3	106	1	PD12_SCHPO
9	112.5	8.3	359	1	PD12_SCHPO
10	111	8.2	643	1	PD44_RAT
11	110.5	8.1	496	1	PD12_DROME
12	110	8.1	107	1	TH12_CORNE
13	110	8.1	122	1	TH12_ORYSA
14	110	8.1	364	1	PD46_MEDSA
15	109	8.0	638	1	PD44_MOUSE
16	108.5	8.0	105	1	TH11_SYNV3
17	108.5	8.0	139	1	TH12_ECOLI
18	108.5	8.0	359	1	TIGA_ASPNG
19	108.5	8.0	645	1	PD44_HUMAN
20	108	7.9	102	1	TH10_CHLPS
21	106	7.8	102	1	TH10_CHLPS
22	106	7.8	102	1	TH10_MYCGE
23	106	7.8	108	1	TH10_BUCAP
24	106	7.8	115	1	TH10_MYCTA
25	105.5	7.8	126	1	TH11_TOBAC
26	105	7.7	509	1	PD11_RABIT
27	105	7.7	510	1	PD11_BOVIN
28	104.5	7.7	119	1	TH22_BRANA
29	104.5	7.7	458	1	TRXB_MYCLE
30	104.5	7.7	522	1	PD11_YEAST
31	104	7.7	108	1	TH10_BUCAP
32	102.5	7.5	105	1	TH10_TREPA
33	102.5	7.5	118	1	TH55_ARATH

34	102.5	7.5	497	1	BS2_TRYBB	PI2865 trypanosoma
35	102	7.5	106	1	TH11_ANASP	P06544 anabaena sp
36	102	7.5	133	1	THH2_ARATH	O38879 arabidopsis
37	102	7.5	361	1	PD46_ARATH	O22263 arabidopsis
38	102	7.5	369	1	ER38_NEUCR	O92249 neurospora
39	102	7.5	509	1	PD1_RAT	P04785 rattus norv
40	101.5	7.5	119	1	TH44_ARATH	O39239 arabidopsis
41	101	7.4	509	1	PD1_MOUSE	P09103 mus musculu
42	100	7.4	104	1	TH13_DICDI	P29447 dictyostell
43	100	7.4	106	1	TH10_PENCH	P34723 penicillium
44	99.5	7.3	118	1	THH3_ARATH	O42403 arabidopsis
45	99.5	7.3	175	1	THIM_WHEAT	O92p21 triticum ae

ALIGNMENTS

RESULT 1
TRX1_YEAST
ID TRX1_YEAST STANDARD; PRT; 103 AA.
AC P22803;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioedoxin I (TR-I)
GN TRX1 OR YGR209C OR G7746.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91107668; PubMed=1988444;
RA Gan Z.-R.;
RT "yeast thioedoxin genes.";
RL J. Biol. Chem. 266:1692-1696(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91225027; PubMed=2026619;
RA Muller E.G.D.;
RT "Thioedoxin deficiency in yeast prolongs S phase and shortens the G1 interval of the cell cycle.";
RL J. Biol. Chem. 266:9194-9202(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c;
RA Guerreiro P., Barreiros T., Soares H., Cyrne L., Maia e Silva A., Rodrigues-Pousada C.;
RT "Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII reveals 12 ORFs, including CCT, ADE3 and TR-I genes, homologues of the yeast PMT and EF1c genes, of the human and bacterial electron-transferring flavoproteins (beta-chain) and of the Escherichia coli phosphoserine phosphohydrolase, and five new ORFs.";
RL Yeast 12:273-280(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=GRF88;
RA Song J.M., Cheung E., Rabinowitz J.C.;
RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-12.
RX STRAIN=YN 295;
RA MEDLINE=95050519; PubMed=7961686;
RT Chae H.-Z., Chung S.J., Rhee S.G.;
RT "Thioedoxin-dependent peroxide reductase from yeast.";
RL J. Biol. Chem. 269:27670-27678(1994).
RN [6]
RP SEQUENCE OF 26-42.
RX MEDLINE=72104597; PubMed=4333602;
RA Hall D.E., Baldesten A., Holmgren A., Reichard P.;
RT "The covalently-bound flavin of hepatic monoamine oxidase. 2. Identification and properties of cysteinyl riboflavin.";

```

RL Eur. J. Biochem. 24:328-335(1971).
CC -!- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62648; AAA35178.1; -
DR EMBL: M59168; AAA35170.1; -
DR EMBL: Z49133; CAA89002.1; -
DR EMBL: 272994; CAA97236.1; -
DR EMBL: U40843; CAA85584.1; -
DR PIR: S15049; TXBY1.
DR PIR: B39847; B39847.
DR PIR: A38669; A38669.
DR HSSP: P10599; IERV.
DR SGD: S0004033; TRX1.
DR InterPro: IPR000063; Thioired.
DR Pfam: PF00085; Thioired; 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFAMS: TIGR01068; thioiredoxin; 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
KW Redox-active center; Electron transport.
FT INIT_MET 0 33 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 30 33 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 103 AA; 11073 MW; 92E6CC4ADF057D31 CRC64;

Query Match 9.1%; Score 124; DB 1; Length 103;
Best Local Similarity 36.7%; Pred. No. 8.6e-05;
Matches 29; Conservative 16; Mismatches 24; Indels 10; Gaps 3;

QY 115 DKRVTVWIFFFANWSDCOSFAPYADLSLKYNCTGLNFGKVDVGRYTDVSTRYKYSTSP 174
Db ||| :||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
18 DKLV--VVDFFATWCGPKMIAPMKFAEQYSDAA--FYKLDVDEVDVAQRAEVSS-- 71

QY 175 LTKQLPTLTLFQGGKAMR 193
Db :||| :||| |
72 ---MPTLIFYKGGKEVTR 86

RESULT 2
THIO_BORBU STANDARD; PRT; 117 AA.
AC O51088;
15-DEC-1998 (Rel. 37, Created)
15-DEC-1998 (Rel. 37, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioiredoxin (TRX).
GN TRXA OR BB0061.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B11;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kierlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT 'Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.';

```

```

RL Nature 390:580-586(1997).
CC -!- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001119; AAC66447.1; -
DR HSSP: P10599; IERV.
DR TIGR: BB0061; -
DR InterPro: IPR000063; Thioired.
DR Pfam: PF00085; Thioired; 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFAMS: TIGR01068; thioiredoxin; 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
KW Redox-active center; Electron transport; Complete proteome.
FT DISULFID 40 43 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 117 AA; 13467 MW; B7CB06F3D316A9D9 CRC64;

Query Match 9.0%; Score 122.5; DB 1; Length 117;
Best Local Similarity 30.3%; Pred. No. 0.00014;
Matches 37; Conservative 22; Mismatches 40; Indels 23; Gaps 6;

QY 100 IKYFNDKTIDEELERKRVTWIVEFFANWSDCOSFAPYADLSLKYNCTGLNFGKVDVG 159
Db :||| :||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12 VKVFDYKNDKEWSFRGDRPA-IIDFYANWCGPKMLSPIFEKLSKYE-NSIDFYKVDTD 69

QY 160 RYTDVSTRYKYSTSPLTQLPTLTLFQGGKAMRRPOIDKKGRAVSWTFSE---ENVIR 215
Db :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
70 KEQDISAIGV-----QSLPTILFI-----PVDGAPK-VSVGFLQEDAFENIK 112

QY 216 EF 217
Db :|
113 DF 114

RESULT 3
TRX2_YEAST STANDARD; PRT; 102 AA.
AC P2217;
01-AUG-1991 (Rel. 19, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioiredoxin II (TR-II).
GN TRX2 OR YLR043C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91107668; PubMed=1988444;
RA Gan Z.-R.;
RT 'Yeast thioiredoxin genes.';
RL J. Biol. Chem. 266:1692-1696(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91225027; PubMed=2026619;
RA Muller E.G.D.;
RT 'Thioiredoxin deficiency in yeast prolongs S phase and shortens the G1
RT interval of the cell cycle.';
RL J. Biol. Chem. 266:9194-9202(1991).
RN [3]
RP SEQUENCE FROM N.A.
RA Koetter P., Rose M., Entian K.D.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

```

```

RN RP SEQUENCE OF 1-12.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fullon L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- FUNCTION: NOT KNOWN.
CC -!- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
CC interchain disulfide bonds in proteins to form the native
CC structures.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
CC -!- SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L15188; AAM22024.1; -
DR PIR; S44756; S4756.
DR HSP; P07237; IMEK.
DR MEROPS; C17.002; -.
DR WormPep; C14B9.2; CE00073.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 3.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRFAMS; TIGR01126; pdi_dom; 3.
DR TIGRFAMS; TIGR01130; ER_PDI_fam; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR PROSITE; PS00194; THIOREDOXIN; 3.
DR Hypothetical protein; Isomerase; Redox-active center;
KW Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 618 PROBABLE PROTEIN DISULFIDE ISOMERASE A4.
FT DISULFID 65 68 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 176 179 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 529 532 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 615 618 PREVENT SECRETION FROM ER.
SQ SEQUENCE 618 AA; 69797 MW; 1DC0207A71444220 CRC64;

Query Match 9.0%; Score 122; DB 1; Length 102;
Best Local Similarity 34.1%; Pred. No. 0.00013;
Matches 29; Conservative 17; Mismatches 29; Indels 10; Gaps 3;

QY 109 DEELERDKRYTWIVFEFANNSDCOSFAPYADLSKYNCTGLNFGKVDVGRYTDVSTRY 168
DB 11 DSAIQDKLV--VDVFATWCGPKMTAPMIEKFESEQ--PQADFYKLDVDELGDVAQKN 66

QY 169 KVSPTSPLTKOLPTLILFQGGKEAMR 193
DB 67 EVSA-----MPTLLLFKNCKEVAK 85

RESULT 4
PDA4_CAEEL
ID PDA4_CAEEL STANDARD; PRT; 618 AA.
AC P34329;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable protein disulfide isomerase A4 precursor (EC 5.3.4.1) (Exp-72
DE homolog).
GN C14B9.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]

```

```

Db 243 AGIIKYMTDOSKPAKKLPKLD 265

RESULT 5
THIO_CAEEL
AC THIO_CAEEL STANDARD; PRT; 115 AA.
ID Q09433.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable thioredoxin B0228.5.
GN B0228.5.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC Rhabditidae; Peleoderinae; Caenorhabditis.
CC NCBI_TaxID=6239;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RA Leimbach D., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D45892; BAA08305.1;
CC DR HSSP; P10599; ITRV.
CC DR InterPro; IPR000063; ThioRed.
CC DR Pfam; PF000085; ThioRed; 1.
CC DR PRINTS; PR00421; THIOREDOXIN.
CC DR PROSITE; PS00194; THIOREDOXIN; 1.
CC KW Redox-active center; Electron transport.
CC FT DISULFID 33
CC FT SEQUENCE 127 AA; 13716 MW; 9F8A02C88EA492EB CRC64;
CC
CC Query Match 8.5%; Score 116; DB 1; Length 127;
CC Best Local Similarity 25.0%; Pred. No. 0.00063;
CC Matches 28; Conservative 27; Mismatches 35; Indels 22; Gaps 4;
CC
CC QY 121 IVFEFFANWSNDCQSFAPYADLSLYKNTCTG-LANFGKVDVGRYTDVSTRYKYSTPLTKQL 179
CC DB 25 VADFIADWCGPCAKAIAPYAFQAKTFSIPNLFAPKINVDVSVQVQVAHYRVA-----M 78
CC QY 180 PTLILFQGGKAMRRPQIDKKGRAVSWTFSEENVIREFNELNYQRAKKLSK 231
CC DB 79 PTFLEFKNGKQV-----AVNGSV---MIOGADVNSLRAAAEKMGK 115

RESULT 7
THIO_MYCPN
AC THIO_MYCPN STANDARD; PRT; 102 AA.
ID P75512.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioredoxin (TRX).
GN TRXA OR TRX OR MPN263 OR MP570.
OS Mycoplasma pneumoniae
CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CC NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Ben-Menachem G., Himmelreich R., Herrmann R., Aharonowitz Y.,
RA Rotten S.;
RT "The thioredoxin reductase system of mycoplasmas.";
RL Microbiology 143:1933-1940(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U51987; AAC45450.1; -;
 DR EMBL; AE000056; AAB96218.1; -;
 DR HSP; P10599; IERV.
 DR InterPro; IPR000063; Thioered.
 DR Pfam; PF00085; thioered; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRFAMs; TIGR01068; thioredoxin; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport; Complete proteome.
 FT DISULFID 30 33 REDOX-ACTIVE (BY SIMILARITY).
 SQ SEQUENCE 102 AA; 11215 MW; 0D17B97E976FC144 CRC64;
 Query Match 8.4%; Score 114; DB 1; Length 102;
 Best Local Similarity 29.5%; Pred. No. 0.00073;
 Matches 26; Conservative 21; Mismatches 31; Indels 10; Gaps 3;
 QY 106 KTIDEELERDKRVTWIVFFANWSDQSFAPYADLSLKYNCTGLNFKGVYDGYRDTYS 165
 Db 9 KQELGFASNNKV--IIDFWAECPCPKITGFEFAKAASEVSTA--FAKVNVDQTDIA 64
 QY 166 TRYKVSPLTKQLPTLILFOGGKEAMR 193
 Db 65 AAYKITS-----LPTIVLFKGGQKHR 86
 RESULT 8
 THIO_COPCM STANDARD; PRT; 106 AA.
 ID Q9UW02, 2001 (Rel. 40, Created)
 AC Q9UW02, 2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Thioredoxin (Allergen Cop c 2).
 OS Coprinus comatus (Shaggy mane).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Agaricales; Agaricaceae; Coprinus.
 NCBI_TaxID=56187;
 [1]
 SEQUENCE FROM N.A.
 RA Brander K.A., Cramer R., Schuermann P., Pichler W.J., Helbling A.;
 RA "Coprinus thioredoxin as inhalative allergen and crossreactive human
 RA autoantigen".;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Participates in various redox reactions through the
 CC reversible oxidation of its active center dithiol to a disulfide
 CC and catalyzes dithiol-disulfide exchange reactions.
 CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ242791; CAB52130.1; -;
 DR HSP; P80028; ITOF.
 DR InterPro; IPR000063; Thioered.
 DR Pfam; PF00085; thioered; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRFAMs; TIGR01068; thioredoxin; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 DR Redox-active center; Electron transport; Allergen.
 KW

FT DISULFID 30 33 REDOX-ACTIVE (BY SIMILARITY).
 SQ SEQUENCE 106 AA; 11772 MW; 05A2155B210E8C69 CRC64;
 Query Match 8.3%; Score 113; DB 1; Length 106;
 Best Local Similarity 30.0%; Pred. No. 0.00095;
 Matches 21; Conservative 17; Mismatches 26; Indels 6; Gaps 1;
 QY 121 IVEFFANWSDQSFAPYADLSLKYNCTGLNFKGVYDGYRDTYSKYKVSPLTKQLP 180
 Db 22 IIDFWATWCGPCRVPIFEKFSKYANNIVFAKVVDVDTASDISEAKI-----RAMP 75
 QY 181 TLILFOGGKE 190
 Db 76 TFQVYKGGQK 85
 RESULT 9
 PD12_SCHPO STANDARD; PRT; 359 AA.
 ID PD12_SCHPO
 AC 013811;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Putative protein disulfide isomerase C17H9.14c precursor (EC 5.3.4.1).
 GN SPAC17H9.14c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OC NCBI_TaxID=4896;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
 RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RL "The genome sequence of Schizosaccharomyces pombe".;
 CC Nature 415:871-880(2002).
 CC -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Achavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.M., Basu P.V., Berman B.P., Bhandari D., Beasley E.M.,
RA Beeson K.I., Benos P., Bhandari D., Bhandari D., Bhandari S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Dey A.D., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin C.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacht J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RT Science 287:2185-2195(2000).
RL
CC -!- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
CC DISULFIDE BONDS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
CC interchain disulfide bonds in proteins to form the native
CC structures.
CC
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (Potential).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL HEAD AND BODY TISSUES.
CC -!- DEVELOPMENTAL STAGE: UBIQUITOUSLY EXPRESSED DURING DEVELOPMENT.
CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18973; AAA86480.1;
CC EMBL; AE003532; AAF49659.1;
CC HSP; P07237; IMEX.
CC FlyBase; FBgn014002; pdi.
CC InterPro; IPR000886; ER_target.
CC InterPro; IPR000063; ThioRed.
CC Pfam; PF00085; ThioRed; 2.
CC PRINTS; PR00421; THIOREDOXIN.
CC TIGRFAMs; TIGR01126; pdi_dom; 2.
CC TIGRFAMs; TIGR01130; ER_PDI_fam; 1.
CC PROSITE; PS00014; THIOREDOXIN; 2.
CC PROSITE; PS00194; THIOREDOXIN; 2.
CC Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
CC SIGNAL 1 18 POTENTIAL.
CC CHAIN 19 496 PROTEIN DISULFIDE ISOMERASE.
CC DISULFID 56 59 REDOX-ACTIVE (BY SIMILARITY).
CC DISULFID 397 400 REDOX-ACTIVE (BY SIMILARITY).
CC DOMAIN 482 496 POLY-GLU.
CC SITE 493 496 PREVENT SECRETION FROM ER (POTENTIAL).
CC SEQUENCE 496 AA; 55781 MW; EB6E04C42167A81 CRC64;

Query Match 8.1%; Score 110.5; DB 1; Length 496;
Best Local Similarity 24.0%; Pred. No. 0.012;
Matches 36; Conservative 25; Mismatches 56; Indels 33; Gaps 5;
QY 110 EELERDKRWIWEFFANNNSDCQSFAPYADLSKY-NCTGLNFGKVDVGRYTDVSTRY 168
DB 378 ESVALDKSKSVLYEFAPWCGHCKQLAPIYDQLAEKVKDNEIVIAKMD-----STAN 430
QY 169 KYVTSPLTKQLPLTLIFQGGKEAMRRPQIDKKGRAVSWTFSEENVIREFNLNELYQRAKK 228
DB 431 ELESIKIS-SFTPIKYFR-----KEDNKKVIDENLDRDLDDFVK 467
QY 229 LSKAGDNIEEPQVASTPTVSGENKKDK 258
DB 468 FLDANGEVADSEPVETEE---EEEAPKKDE 495
RESULT 12
THI2_CORNE
ID THI2_CORNE STANDARD; PRT; 107 AA.
AC P07887;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ThioRedoxin C-2.
OS Corynebacterium nephridii.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID:1722;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308211; PubMed=3040729;
RA Lim C.-J., Fuchs J.A., McFarlan S.C., Hogenkamp H.P.C.;
RT "Cloning, expression, and nucleotide sequence of a gene encoding a
RL second thioRedoxin from Corynebacterium nephridii.";
RN [2]
RP J. Biol. Chem. 262:12114-12119(1987).
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89137116; PubMed=2917572;
RA McFarlan S.C., Hogenkamp H.P.C., Eccleston E.D., Howard J.B.,
RA Fuchs J.A.;
RT "Purification, characterization and revised amino acid sequence of a
RL second thioRedoxin from Corynebacterium nephridii.";
CC Eur. J. Biochem. 179:389-398(1989).
CC -!- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14630; CAA32779.1;
CC EMBL; J02801; AAA23305.1; ALT_INIT.
CC FIR; S02802; S02802.
CC HSP; P00274; 2TRX.
CC InterPro; IPR000063; ThioRed.
CC Pfam; PF00085; ThioRed; 1.
CC PRINTS; PR00421; THIOREDOXIN.
CC TIGRFAMs; TIGR01068; ThioRedoxin; 1.
CC PROSITE; PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport; Multigene family.
CC INIT MET 0 0
CC DISULFID 32 35 REDOX-ACTIVE.
CC SEQUENCE 107 AA; 11585 MW; C21EB09648FAFALC CRC64;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 12, 2003, 09:47:42 ; Search time 185 Seconds
(without alignments)
3140.625 Million cell updates/sec

Title: US-09-954-846-2

Perfect score: 1359

Sequence: 1 MAVLAPLIAIVSVPLSRW.....EQPVASTPTTVSDGENKKDK 258

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

arched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09954846/runat_09072003_102144_12148/app_query.fasta_1.455
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09954846 -CGN_1_1_200 -runat_09072003_102144_12148 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1359	100.0	774	21	AAAL5981	Human protein clon
2	1359	100.0	1527	21	AAAL5991	Human protein clon
3	1359	100.0	1598	21	AAF16316	Human prostate can
4	1326	97.6	1618	20	AAAX5251	Protein PRO270 CDN
5	1326	97.6	1620	22	AAAS45940	Human CCI-31 prote
6	1326	97.6	1730	22	ABA09248	Human CGI-31 prote
7	1326	97.3	1699	19	AAV34294	Human secreted pro
8	1313	96.6	1621	22	AAF72409	Human PRO270 CDNA
9	1313	96.6	1652	19	AAV34316	Human secreted pro
10	1308	96.2	1659	24	ABL49614	Prostaglandin E2 (
11	1299	95.6	1631	22	AAF93762	Human CDNA encodin
12	984	73.1	774	21	AAZ42527	Human 5' EST isola
13	956	70.3	726	22	AAF93955	Primer specific fo
14	925.5	68.1	2030	23	AAAS8076	DNA encoding novel
15	622	45.8	439	21	AAA41964	Human secreted exp
16	620	45.6	461	21	AAAC01008	Human secreted pro
17	579.5	42.6	3369	22	ABA15532	Human nervous syst
18	579.5	42.6	3369	22	ABA15533	Human nervous syst
19	514	37.8	444	20	AA41205	Human secreted pro
20	498	36.6	2297	22	ABA14531	Human nervous syst
21	476.5	35.1	919	23	ABL17145	Drosophila melanog
22	476.5	35.1	2919	23	ABL17145	Drosophila melanog
23	447	32.9	430	20	AA41202	Human secreted pro
24	431.5	31.8	5423	23	AAAS8062	DNA encoding novel
25	431.5	31.8	5423	24	ABL68483	Kidney cancer rela
26	430	31.6	452	20	AA41203	Human secreted pro
27	424	31.2	307	23	AAAS8053	DNA encoding novel
28	422	31.1	469	21	AAZ43018	Human 5' EST isola
29	415	30.5	307	24	ABK45061	CDNA encoding colo
30	372	27.4	274	21	AAA45185	Human secreted exp
31	353.5	26.0	756	23	ABV21932	Human prostate exp
32	353.5	26.0	756	23	ABV27764	Human prostate exp
33	336	24.7	1895	23	AAAS8073	DNA encoding novel
34	323	23.8	1409	22	ABA15530	Human nervous syst
35	280	20.6	531	21	AAZ43016	Human 5' EST isola
36	269	19.8	153	22	AAAL19490	Human breast cance
37	269	19.8	433	22	AALO9103	Human prostate exp
38	232	17.1	370	23	ABV02989	Human nervous syst
39	175	12.9	367	22	ABAL1541	Human prostate exp
40	163.5	12.0	375	23	ABV33303	Human prostate exp
41	163.5	12.0	375	23	ABV42226	Human prostate exp
42	163.5	12.0	389	23	ABV12158	Human prostate exp
43	130.5	9.6	3550	9	AA81540	Sequence of pTrx-2
44	130.5	9.6	3550	13	AAQ31623	pTrx-2 containing
45	129.5	9.5	3552	11	AAQ05875	pTrx-2 plasmid seq

ALIGNMENTS

RESULT 1

AAAL5981
ID AAAL5981 standard; CDNA; 774 BP.

XX AAAL5981;

AC AAAL5981;

DT 12-JUN-2000 (first entry)

DE Human protein clone HP10392 coding sequence.

XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW cytokine production; cell proliferation; cell differentiation;
KW immune deficiency; infectious disease; autoimmune disorder; asthma;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;

coagulation disorder; myocardial infarction; inflammatory condition; septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour; nephritis; therapy; ss.

Homo sapiens.

WO200005367-A2.

03-FEB-2000.

22-JUL-1999: 99WO-JP03929.

24-JUL-1998; 98JP-0208820.

07 AUG 1998, 98JF-022410J.
25-AUG-1998: 98JP-0238116:

03-SEP-1998; 98JF-0234730;
29-SEP-1998; 98JP-0275505;

(SAGA) SAGAMI CHEM RES CENT

(FROTH) PROTEGENE INC.

'T PITY 'S OJBY

WFL, 2000-102034/10.
P-PSDB; AAY94893.

Novel human protein

multiple sclerosis, Alzheimer's disease, and

Claim 3: Page 328

This sequence encodes a human protein of the invention, which has hydrophobic domains. The DNA sequences can be used as a probe or as a genetic marker. The protein can also be used as a marker, and to identify potential genetic disorders. The DNA and protein can also be used as nutritional sources or supplements. The protein exhibits cytokine, cell proliferation, cell differentiation activities and induces production of other cytokines in certain cell populations. The protein also exhibits immune stimulating or immune suppressing activity. It can be used in the treatment of various immune deficiencies and disorders, and to treat infectious diseases caused by viral, bacterial, fungal or other infections. The protein is also used for treating autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid arthritis. It is also useful in the treatment of allergic reactions and conditions such as asthma, and in immune suppression after organ transplantation. The protein is useful in regulation of haematopoiesis and consequently in the treatment of myeloid or lymphoid cell deficiencies. It is also used in compositions for tissue growth or

regeneration. The protein is also used in the treatment of osteoporosis or osteoarthritis and in the treatment of periodontal disease and other tooth repair processes. The protein is used in the treatment of nervous system disorders such as Alzheimer's disease, Parkinson's disease, and Huntington's disease. They are useful for protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage. They are also used for promoting or inhibiting tissue differentiation. They are also used as contraceptives since they exhibit activin or inhibin related activities and as a fertility inducing therapeutic. They are used for treating various coagulation disorders and in treatment and prevention of conditions resulting from coagulation activities e.g. myocardial infarction or stroke. They also acts as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. They are used to treat inflammatory conditions such as septic shock, sepsis, ischaemia reperfusion injury, arthritis, and nephritis. They can be used to prevent tumours.

Sequence 774 BP: 200 A: 200 C: 190 G: 184 T: 0 other:

Alignment Scores:

Alignment scores:	
Pred. No.:	3.41e-155
Length:	774

Score:	1359.00
Matches:	

Percent Similarity: 100.00%


```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
```

US-09-954-846-2 (1-258) x AA15981 (1-774)

Qy 1 MetAlaValLeuAlaProLeuIleAlaL

1	ATGGCGGTCTTGGCACCTCTAAATTGCTCTCGTATTCTGGTCCGCGACTTTCACGATGG	60
21	LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuLeuValArg	40
61	CTGCCCAACCTACTACTCTCTCTGGCCCTGCTCTCTGCTGCCTTCTACTCTGAGG	120
41	LysLeuProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp	60
121	AAACTGGCGCGTCTGCCACGGTCTGCCACCACGCGAAGACGGTAACCCGTGTGAC	180
61	PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn	80
181	TTTGACTGGAGAGAAGTGGAGATCCTGATGTTCTCAGTGCCATGTGATGATGAAGAC	240
81	ArgArgSerMetPheLeuMetThrCysLysProProLeuTyrMetGlyProGluTyrIle	100
241	CGCAGATCCATGTTCTCTGATGACGTGCAAACCCCCCTATATATGGGCCCTGAGTATATC	300
101	LysTyrPheAsnAspLysThrIleAspGluLeuGluArgAspLysArgValThrTrp	120
301	AAGTACTTCAATGATATAAAGCATTTGATGAGGAAGTAGAACGGGACAGAGGGTCACTGG	360
121	IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAla	140
361	ATTGTGGAGTCTTTGGCAATTGGCTTAATGACTGCCAATCATTTGCCCTCATCTATGCT	420
141	AspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPheGlyLysValAspValGlyArg	160
421	GACCTCTCCCTTAAATACAACTGTACAGGCTAAATTTTGGAGGTGGATGTTGGACCG	480
161	TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro	180
481	TATACTGATCTTAGTACGCGGTACAAAGTGGAGCATCACCCCTCACCAAGCAACTCCCT	540
181	ThrLeuIleLeuPheGlnGlyLysGluAlaMetArgArgProGlnIleAspLysLys	200
541	ACCCTGATCTGTTCCTCAAGTGGCAAGGAGCAATGCGCGGCCACAGATTGACAAGAAA	600
201	GlyArgAlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn	220
601	GGACGGCTCTCATGGACCTTCTCTGAGGAGAATGTGATCCGAGAATTTAACTTAAAT	660
221	GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGln	240
661	GAGCTATACAGCGGGCCAGAAACTATCAAAGCTGGAGACAATATCCCTGAGGAGCAG	720
241	ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys	258
721	CTGTGGCTTCAACCCACCCACAGCTGTAGATGGGAAAAACAAGAGCATAAA	774

RESULT 2

~~AAA~~15991

ID \ AAA15991 standard; cDNA; 1527 BP.

XX

AC ~~AAA15991;~~
vv

XX
DT
13 → TIN-2000 (first entries)

DT JUN-2000 (first entry)
XX

XX
DE Human protein clone HP10392 full length coding sequence

~~human protein clone AF0392 rat tenascin coding sequence.~~

Human protein; hydrophobic domain; nutritional source; haematopoiesis; cytokine production; cell proliferation; cell differentiation; immune deficiency; infectious disease; autoimmune disorder; asthma; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; allergic reaction; osteoporosis; osteoarthritis; periodontal disease.

3

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OR nucleic - nucleic search, using sw model

Run on: July 12, 2003, 08:31:01 ; Search time 269 Seconds
(without alignments)
13018.057 Million cell updates/sec

Title: US-09-954-846-4

Perfect score: 1555

Sequence: 1 AGGGGAGCGGGCGGAGACC.....TAATAAAGATTCGGATTA 1555

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by change to have a score greater than or equal to the score of the results being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1520	97.7	1527	21	AAAL5991 Human protein clon
2	1511	97.2	1598	21	AAAL5991 Human prostate can
3	1429	91.9	1730	22	ABA09248 Human CGI-33 prote
4	1381.6	88.8	1699	19	AAV34294 Human secreted pro
5	1370.2	88.1	1631	22	AAV33762 Human cDNA encoding
6	1370	88.1	1620	22	AAAS45940 Human DNA encoding
7	1368	88.0	1618	20	AAAS52251 Protein PRO270 cDN
8	1361.8	87.6	1652	19	AAV34316 Human secreted pro
9	1361.2	87.5	1659	24	ABL49614 Prostaglandin E2 (

10	1359	87.4	1621	22	AAF72409 Human PRO270 cDNA.
11	868	55.8	2297	22	ABAL5531 Human nervous syst
12	868	55.8	3369	22	ABAL5532 Human nervous syst
13	868	55.8	3369	22	ABAL5533 Human nervous syst
14	774	49.8	774	21	AAAL5981 Human protein clon
15	613	39.4	2030	23	AAAS8076 DNA encoding novel
16	521	33.5	774	21	AAZ43527 Human 5' EST isola
17	463	29.8	726	22	AAAF9355 Primer specific fo
18	441.2	28.4	521	16	AAAT22413 Human gene signatu
19	348.4	22.4	439	21	AAAA1964 Human secreted exp
20	308	19.8	469	21	AAZ43018 Human 5' EST isola
21	305.6	19.7	452	20	AAAX41203 Human secreted pro
22	305.2	19.6	461	21	AAC01008 Human secreted pro
23	287.2	18.5	430	20	AAAX41202 Human secreted pro
24	261	16.8	307	23	AAAS8053 DNA encoding novel
25	261	16.8	444	20	AAAX41205 Human secreted pro
26	258.4	16.6	307	24	ABK45061 cDNA encoding colo
27	252.2	16.2	5423	23	AAAS8062 DNA encoding novel
28	252.2	16.2	5423	24	ABK45062 Kidney cancer rela
29	246	15.8	1895	23	AAAS8073 DNA encoding novel
30	230.6	14.8	531	21	AAZ43016 Human 5' EST isola
31	221	14.2	756	23	ABV21932 Human prostate exp
32	221	14.2	756	23	ABV21932 Human prostate exp
33	200.4	12.9	274	21	AAA45185 Human secreted exp
34	144.6	9.3	153	22	AAAL19490 Human breast cance
35	144.6	9.3	433	22	AAAL09103 Human breast cance
36	141.8	9.1	370	23	ABV02989 Human prostate exp
37	110.2	7.1	1409	22	ABAL5530 Human nervous syst
38	95.8	6.2	2919	23	ABAL17144 Prosopilla melanog
39	89.2	5.9	919	23	ABAL17145 Prosopilla melanog
40	89.2	5.7	367	22	ABAL1541 Human nervous syst
41	84.4	5.4	375	23	ABV33303 Human prostate exp
42	84.4	5.4	375	23	ABV33303 Human prostate exp
43	84.4	5.4	389	23	ABV12158 Human prostate exp
44	50	3.2	50	20	AAAS2405 Probe used to iso
45	50	3.2	50	22	AAF72563 Human PRO polypt

ALIGNMENTS

RESULT 1

AAAL5991

ID AAAL5991 standard; cDNA; 1527 BP.

XX AAAL5991;

AC AAAL5991;

DT 12-JUN-2000 (first entry)

XX

XX Human protein clone HP10392 full length coding sequence.

DE

XX

KW

KW

KW

KW

KW

KW

KW

KW

KW

OS

XX

XX

PN

XX

PD

XX

XX

XX

PR

PR

Human protein; hydrophobic domain; nutritional source; haematopoiesis; cytokine production; cell proliferation; cell differentiation; immune deficiency; infectious disease; autoimmune disorder; asthma; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; allergic reaction; osteoporosis; osteoarthritis; periodontal disease; nervous system disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury; systemic cytokine damage; tissue differentiation; contraceptive; stroke; coagulation disorder; myocardial infarction; inflammatory condition; septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour; nephritis; therapy; ss.

Homo sapiens.

WO200005367-A2.

03-FEB-2000.

22-JUL-1999; 99WO-JP03929.

24-JUL-1998; 98JP-0208820.

07-AUG-1998; 98JP-0224105.

25-AUG-1998; 98JP-0238116.
 09-SEP-1998; 98JP-0254736.
 29-SEP-1998; 98JP-0275505.
 (SAGA) SAGAMI CHEM RES CENT.
 (PROT-) PROTEGENE INC.
 Kato S, Kimura T;
 WPI; 2000-182694/16.
 P-PSDB; AAY94893.
 Novel human proteins having hydrophobic domains useful for treating osteoporosis, Alzheimer's disease, Parkinson's disease, asthma, multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 Claim 4; Page 341-343; 351pp; English.
 This sequence encodes a human protein of the invention, which has hydrophobic domains. The DNA sequences can be used as a probe or as a genetic marker. The protein can also be used as a marker, and to identify potential genetic disorders. The DNA and protein can also be used as nutritional sources or supplements. The protein exhibits cytokine, cell proliferation, cell differentiation activities and induces production of other cytokines in certain cell populations. The protein also exhibits immune stimulating or immune suppressing activity. It can be used in the treatment of various immune deficiencies and disorders, and to treat infectious diseases caused by viral, bacterial, fungal or other infections. The protein is also used for treating autoimmune disorders such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid arthritis. It is also useful in the treatment of allergic reactions and conditions such as asthma, and in immune suppression after organ transplantation. The protein is useful in regulation of haematopoiesis and consequently in the treatment of myeloid or lymphoid cell deficiencies. It is also used in compositions for tissue growth or regeneration. The protein is also used in the treatment of osteoporosis or osteoarthritis and in the treatment of periodontal disease and other tooth repair processes. The protein is used in the treatment of nervous system disorders such as Alzheimer's disease, Parkinson's disease, and Huntington's disease. They are useful for protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage. They are also used for promoting or inhibiting tissue differentiation. They are also used as contraceptives since they exhibit activin or inhibin related activities and as a fertility inducing therapeutic. They are used for treating various coagulation disorders and in treatment and prevention of conditions resulting from coagulation activities e.g. myocardial infarction or stroke. They also act as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. They are used to treat inflammatory conditions such as septic shock, sepsis, ischaemia reperfusion injury, arthritis, and nephritis. They can be used to prevent tumours.

181 CGCGAAGACGGTAACCCGTTGACCTTGAGTGGAGAGAAAGTGAGATCTGATGTTTCTC 240
 275 AGTGCCATTGTGATGATGAAGAACCCGAGATCCATGTTCTCTGATGACGTGCAAAACCCCC 334
 241 AGTGCCATTGTGATGATGAAGAACCCGAGATCCATGTTCTCTGATGACGTGCAAAACCCCC 300
 335 CTATATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCACTTGTAGGAGAACTA 394
 301 CTATATATGGGCCCTGAGTATATCAAGTACTTCAATGATAAAACCACTTGTAGGAGAACTA 360
 395 GAACGGGACACAGAGGCTCACTTGGATTGTGGAGTCTTTGGCCAAATGGTCTTAATGACTGC 454
 361 GAACGGGACACAGAGGCTCACTTGGATTGTGGAGTCTTTGGCCAAATGGTCTTAATGACTGC 420
 455 CAATCATTTGGCCCTCATCTATCTGACCTCTCCCTTAAATACAACTGTACAGGCTAAAT 514
 421 CAATCATTTGGCCCTCATCTATCTGACCTCTCCCTTAAATACAACTGTACAGGCTAAAT 480
 515 TTTGGGAAGGTGATGTTGGACGCTATCTAGTGTAGTACGGGTACAAAGTGACGACACA 574
 481 TTTGGGAAGGTGATGTTGGACGCTATCTAGTGTAGTACGGGTACAAAGTGACGACACA 540
 575 TCACCCCTCACCAAGCAACTCCCTACCTGATCTTCCAAAGGTGGCAAGGAGGCAATG 634
 541 TCACCCCTCACCAAGCAACTCCCTACCTGATCTTCCAAAGGTGGCAAGGAGGCAATG 600
 635 CGCGGGCCACAGATTGACAAAGAGGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAT 694
 601 CGCGGGCCACAGATTGACAAAGAGGCGGCTGTCTCATGGACCTTCTCTGAGGAGAAT 660
 695 GTGATCCGAGAAATTTAACTTAATGAGCTATACAGCGGGCCCAAGAACTATCAAAAGGT 754
 661 GTGATCCGAGAAATTTAACTTAATGAGCTATACAGCGGGCCCAAGAACTATCAAAAGGT 720
 755 GGAGACAATATCCCTGAGGAGGAGCTGTGGCTTCAACCCCGACACACGTGTCAGATGG 814
 721 GGAGACAATATCCCTGAGGAGGAGCTGTGGCTTCAACCCCGACACACGTGTCAGATGG 780
 815 GAAACACAGAGGATAAATAAGATCCTCACTTTGGCAGTCTTCTCTCTCTCAATTC 874
 781 GAAACACAGAGGATAAATAAGATCCTCACTTTGGCAGTCTTCTCTCTCTCAATTC 840
 875 AGGCTCTTCCATACACACAGGCTGAGGCTGAGGCTGAGGCTTCTCTCTCTCTCTCTCT 934
 841 AGGCTCTTCCATACACACAGGCTGAGGCTGAGGCTGAGGCTTCTCTCTCTCTCTCTCT 900
 935 CTGTGACTGGTGGGCGGACATGTCAGCTTCTGATTTTAAAGAGGCACTTCTAGGGAATGTC 994
 901 CTGTGACTGGTGGGCGGACATGTCAGCTTCTGATTTTAAAGAGGCACTTCTAGGGAATGTC 960
 995 AGGCACCTACAGGAAGGCTGCCATGCTGTGGCCCACTGTTTCTCACTGGAGCAAGAAAGA 1054
 961 AGGCACCTACAGGAAGGCTGCCATGCTGTGGCCCACTGTTTCTCACTGGAGCAAGAAAGA 1020
 1055 GATCTCATAGACGAGGAGGGAATGTTTCCCTCCAGCTTGGGTGAGTGTGTTAACTG 1114
 1021 GATCTCATAGGACGAGGAGGGAATGTTTCCCTCCAGCTTGGGTGAGTGTGTTAACTG 1080
 1115 CTTATCAGCTTATCAGACATCTCCATGCTTCTCCATGAAACTCTGTGGTTTCATCATTC 1174
 1081 CTTATCAGCTTATCAGACATCTCCATGCTTCTCCATGAAACTCTGTGGTTTCATCATTC 1140
 1175 CTTCTTACTGACCTGACAGCTTGGTAGACCTAGATTTAACCTTAAGGTAAAGTGTG 1234
 1141 CTTCTTACTGACCTGACAGCTTGGTAGACCTAGATTTAACCTTAAGGTAAAGTGTG 1200
 1235 GGGTATAGACGCTAAGAAATTTCCCAAGGACTCTTCTGCTTCCCTTAAGCCCTTCTGGCT 1294
 1201 GGGTATAGACGCTAAGAAATTTCCCAAGGACTCTTCTGCTTCCCTTAAGCCCTTCTGGCT 1260
 1295 TCGTTTATGCTTCAATTAAGATATAGCCTAACTTTGTCGCTAGTCTCTAAGGAGAAAC 1354
 1261 TCGTTTATGCTTCAATTAAGATATAGCCTAACTTTGTCGCTAGTCTCTAAGGAGAAAC 1320

Sequence 1527 BP; 382 A; 373 C; 361 G; 411 T; 0 other;

Query Match 97.7%; Score 1520; DB 21; Length 1527;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 35 GCAGTGGCCGTTACGGCGGAAAGATGGCGTCTTGGCACTCTAATGCTCTCGTGAT 94
 1 GCAGTGGCCGTTACGGCGGAAAGATGGCGTCTTGGCACTCTAATGCTCTCGTGAT 60
 95 TCGGTGGCCGACCTTACAGATGGCTGCCCAACCTTACTACCTTCTGTCGGCCCTGCTC 154
 61 TCGGTGGCCGACCTTACAGATGGCTGCCCAACCTTACTACCTTCTGTCGGCCCTGCTC 120
 155 TCTGCTGCTTCTTCTACGTCGTGAGAAACTGCCCGGCTCTGCCAGGCTCTGCCACCCAA 214
 121 TCTGCTGCTTCTTCTACGTCGTGAGAAACTGCCCGGCTCTGCCAGGCTCTGCCACCCAA 180
 215 CGCGAAGACGGTACCGGTGACTTTGACTGGAGAGAAAGTGAGATCTGATGTTTCTC 274

QY 1355 CTTTAAACACAAAGTTTATCATTTAGAGACAATATTGAACACCCCTATTGTTGGG 1414
Db CTTTAAACACAAAGTTTATCATTTAGAGACAATATTGAACACCCCTATTGTTGGG 1380
QY 1415 ATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCTTTGTTGTTAGACTTTGGAGGAGA 1474
Db ATTGAGAGGGGTGAATAGAGGCTTGAGACTTTCTTTGTTGTTAGACTTTGGAGGAGA 1440
QY 1475 AATCCCTCGGACTTTTCACTAACCTCTGACATCTCCACACACCCAGTTGATGCTTCC 1534
Db AATCCCTCGGACTTTTCACTAACCTCTGACATCTCCACACACCCAGTTGATGCTTCC 1500
QY 1535 GTAATAAAAGATTGGGATT 1554
Db GTAATAAAAGATTGGGATT 1520

RESULT 2

AAFI6316

ID AAF16316 standard; cDNA; 1598 BP.

AAFI6316;

13-MAR-2001 (first entry)

Human prostate cancer antigen nucleotide sequence SEQ ID NO:751.

Humap; prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytoskeletal; cardiovascular; immunomodulatory; muscular;
vulnerable; gastrointestinal; nephrotropic; anti-infective; gynaecological;
antibacterial; gene therapy; neural; immune; reproductive; renal;
gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
wound; infectious disease; ss.

Homo sapiens.

WO200055174-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05988.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

(ROSE/) ROSEN C A.

Rosen CA, Ruben SM;

WPI; 2000-587513/55.

P-PSDB; AAB57113.

Prostate cancer associated gene sequences, referred to as prostate
cancer antigens, useful for treatment, prevention, and diagnosis of
disorders such as prostate cancer.

Claim 1; Page 1172; 2338pp; English.

AAFI5566 to AAFI6505 encode the human prostate cancer associated
proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
The prostate cancer antigens can have neuroprotective, cytoskeletal,
cardioactive, immunomodulatory, muscular, vulnerable, gastrointestinal,
nephrotropic, anti-infective, gynaecological and antibacterial activities,
and can be used in gene therapy. The prostate cancer antigen
polynucleotides may be used for detection of prostate cancer, chromosome
identification, as chromosome markers, and for numerous other diagnostic
or research purposes. The prostate cancer antigens may be used to treat
disorders such as neural, immune, muscular, reproductive,
gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
disorders, wounds, and infectious diseases. AAFI6506 to AAFI6514 to
AAB57303 represent sequences used in the exemplification of the present
invention.

XX

SQ Sequence 1598 BP; 437 A; 381 C; 365 G; 414 T; 1 other;

Query Match 97.2%; Score 1511; DB 21; Length 1598;

Best Local Similarity 99.9%; Pred. NO. 0;

Matches 1511; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 42 CCGTTACGGCGCAAAAGATGCGGTCTTGGCACCTCTAATTGCTCTCGTGTATTCGGTGC 101
Db CCGTTACGGCGCAAAAGATGCGGTCTTGGCACCTCTAATTGCTCTCGTGTATTCGGTGC 85
QY 102 CGGACATTTACAGATGCGTCCGCAACCTTACTACTCTCTCGGCGCTCTCTCTGCTG 161
Db CGGACATTTACAGATGCGTCCGCAACCTTACTACTCTCTCGGCGCTCTCTCTGCTG 145
QY 162 CTTTCTACTCTGAGGAAACTGCGCGCTCTGCCACGGTCTGCCACCGTCTGCCACCGAAG 221
Db CTTTCTACTCTGAGGAAACTGCGCGCTCTGCCACGGTCTGCCACCGTCTGCCACCGAAG 205
QY 222 ACGCAACCCGCTGTGACTTTGACTGGAGAGAGTGGAGATCTTGATGTTCTCAGTGCCA 281
Db ACGTAACCCGCTGTGACTTTGACTGGAGAGAGTGGAGATCTTGATGTTCTCAGTGCCA 265
QY 282 TTGTGATGATGAAGAAGCGCAGATCCATGTTCTTGATGACGTGCAAAACCCCTATATA 341
Db TTGTGATGATGAAGAAGCGCAGATCCATGTTCTTGATGACGTGCAAAACCCCTATATA 325
QY 342 TGGCCCTCGAGTATTTCAAGTACTTCAATGATAAAACCATTTGATGAGGAATAGAACGGG 401
Db TGGCCCTCGAGTATTTCAAGTACTTCAATGATAAAACCATTTGATGAGGAATAGAACGGG 385
QY 402 ACAAGAGGTCACATGGATGCTGAGTCTTTGGCAATTTGCTTAATGACGTGCCAATCAT 461
Db ACAAGAGGTCACATGGATGCTGAGTCTTTGGCAATTTGCTTAATGACGTGCCAATCAT 445
QY 462 TTGCCCTATCTATGCTGACCTTCCCTTAATCAACTGCTACAGGGCTAATTTTGGGA 521
Db TTGCCCTATCTATGCTGACCTTCCCTTAATCAACTGCTACAGGGCTAATTTTGGGA 505
QY 522 AGTGGATGTTGGACGCTATACTGATGTAGTAGCGGTACAAAGTGAAGCATCACCC 581
Db AGTGGATGTTGGACGCTATACTGATGTAGTAGCGGTACAAAGTGAAGCATCACCC 565
QY 582 TCACCAAGCACTCCCTACCTGATCTCTTCCAAAGTGGCAAGAGCAATGCGCGGC 641
Db TCACCAAGCACTCCCTACCTGATCTCTTCCAAAGTGGCAAGAGCAATGCGCGGC 625
QY 642 CACAGATTGACAAGAGGCGGCTGTCTCATGACCTTCTCTGAGGAGATGTGATCC 701
Db CACAGATTGACAAGAGGCGGCTGTCTCATGACCTTCTCTGAGGAGATGTGATCC 685
QY 702 GAGAAATTAATTAATGAGCTATACAGCGGGCAAGAAACTATCAAGGCTGGAGACA 761
Db GAGAAATTAATTAATGAGCTATACAGCGGGCAAGAAACTATCAAGGCTGGAGACA 745
QY 762 ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGTCAAGTGGGAAAACA 821
Db ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACACAGTGTCAAGTGGGAAAACA 805
QY 822 AGAAGGATAAATAGATCCTTACCTTTGGCAGTGCTTCTCTCTGTCATTTCCAGGCTCT 881
Db AGAAGGATAAATAGATCCTTACCTTTGGCAGTGCTTCTCTCTGTCATTTCCAGGCTCT 865
QY 882 TTCATTAACACAGCCTGAGGCTGACGCTTTTATTTATGCTTTTCCCTTTGGCTGAC 941
Db TTCATTAACACAGCCTGAGGCTGACGCTTTTATTTATGCTTTTCCCTTTGGCTGAC 925
QY 942 TGGGTGGGCGAGCATGAGCTTCTGATTTTAAAGAGCATCTAGGGAATTTCTAGGCAAC 1001
Db TGGGTGGGCGAGCATGAGCTTCTGATTTTAAAGAGCATCTAGGGAATTTCTAGGCAAC 985
QY 1002 CTACAGGAAGCCTGCCATGCTGTGGCAACTGTTTCACTGGAGCAAGAGATCTCA 1061
Db CTACAGGAAGCCTGCCATGCTGTGGCAACTGTTTCACTGGAGCAAGAGATCTCA 1045

201 GlyArgAlaValSerTrpThrPheSerGluGluAsnValIleArgGluPheAsnLeuAsn 220
 Db GGACGGCGTGTCTCATGGACCTTCTCTGAGGAGAATGTGATCCGAGAAATTTAACTTAAT 718
 221 GluLeuTyrGlnArgAlaLysLysLysLeuSerLysAlaGlyAspAsnIleProGluGln 240
 Db GAGCTTACCAACGCGGCCAAGAACTATCAAGAGCTGGAGACAATATCCCTGAGGAGCAG 778
 241 ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
 Db CCTGTGGCTTCAACCCCAACCAAGCTGTCAGATGGGGAACAGAGATATA 832
 RESULT 2
 US-09-925-300-751
 ; Sequence 751, Application US/09925300
 ; Patent No. US20020151681A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Craig Rosen,
 ; APPLICANT: Steve Ruben
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA101
 ; CURRENT APPLICATION NUMBER: US/09/925,300
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05988
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 1890
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 751
 ; LENGTH: 1598
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-300-751
 Alignment Scores:
 Pred. No.: 4, 26-180 Length: 1598
 Score: 1359.00 Matches: 238
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-09-954-846-2 (1-258) x US-09-925-300-751 (1-1598)
 QY 1 MetaAlaValAlaProLeuIleAlaLeuValTyrSerValProArgLeuSerArgTrp 20
 Db 43 ATGGCGGTCTGGACCTCTAATTGCTCTGCTATTCGGTGGCGGCACTTTCACCATGG 102
 QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
 Db 103 CTCGCCCAACCTTACTACCTTCTGTGCGCCCTGCTCTGCTCTCTCTCTCTCTCTCTGAGG 162
 QY 41 LysLeuProLeuLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysAsp 60
 Db 163 AAACCTGCGCGCGCTCTGCCACGCTTGGCCACCCCAACGCGAAGACGGTAACCCGCTGTGAC 222
 QY 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetLysAsn 80
 Db 223 TTGTACTGGAGAGAGTGGAGATCTCGATTTCTCAGTGCCATTCTGATGATGAAGAAC 282
 QY 81 ArgArgSerMetPheLeuMetThrCysLysProLeuTyrMetGlyProGluTyrIle 100
 Db 293 CGCAGATCCATGTTCTCTGATGACGTGCACACCCCTATATATGCGCCCTGAGTATATC 342
 QY 101 LysTyrPheAsnAspLysThrIleAspGluGluLeuGluArgAspLysArgValThrTrp 120
 Db 343 AAGTACTTCAATGATAAACCATTGATGAGGAACTAGAACGGGACAAAGAGGGTCACTTGG 402
 QY 121 IleValGluPhePheAlaAsnTrpSerAsnAspCysGlnSerPheAlaProIleTyrAla 140
 Db 403 ATTTGGAGGAGTCTTTGGCAATTTGGTCTAATGACTGCAATCATTTTCCCTATCTATGCT 462

QY 141 AspLeuSerLeuLysTyrAsnCysThrGlyLeuAsnPhcGlyLysValaspValGlyArg 160
DB 463 GACCTCTCCCTTAAATACAACTGTACAGGCTAAATTTGGGAAGGTGGATGTTGGACGC 522
QY 161 TyrThrAspValSerThrArgTyrLysValSerThrSerProLeuThrLysGlnLeuPro 180
DB 523 TATACCTGATGTAGTACGGGTACAAGAGGACATACCCCTCACCAGCAACATCCCT 582
QY 181 ThrLeuLeuPheGlnGlyLysGluAlaMetArgProGlnLeuAspLysLys 200
DB 583 ACCCTGATCTGTCTCAAGGTGGCAAGGCAATCGCGGCCACAGATTGACAAGAA 642
QY 201 GlyArgAlaValSerThrPheSerGluGluAsnValleArgGluPheAsnLeuAsn 220
DB 643 GGACGGCTGTCTCATGGACCTTCTCTGAGGAGATGTGATCGGAGAAATTAACCTAAAT 702
221 GluLeuTyrGlnArgAlaLysLysLeuSerLysAlaGlyAspAsnIleProGluGluGln 240
703 GAGCTATACACGGGGCCCAAGAACTATCAAGGCTGGAGACAATATCCCTGAGGAGCAG 762
QY 241 ProValAlaSerThrProThrThrValSerAspGlyGluAsnLysLysAspLys 258
DB 763 CCTGTGGCTTCAACCCCAACACAGTGTACAGTGGGGAACCAAGAGGATAAA 816

RESULT 3

US-09-905-291A-206

Sequence 206, Application US/09905291A

Patent No. US2002016037A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kajiav, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Paul, James
APPLICANT: Paoli, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE OF INVENTION: clds Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/905,291A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-07-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 206
LENGTH: 1620
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (973)
OTHER INFORMATION: a, t, c or g
NAME/KEY: modified_base
LOCATION: (977)
OTHER INFORMATION: a, t, c or g
NAME/KEY: modified_base
LOCATION: (996)
OTHER INFORMATION: a, t, c or g
NAME/KEY: modified_base
LOCATION: (1003)
OTHER INFORMATION: a, t, c or g
US-09-905-291A-206

Alignment Scores:

Pred. No.: 1.84e-175 Length: 1620
Score: 1326.00 Matches: 257
Percent Similarity: 87.16% Conservative: 1
Best Local Similarity: 86.82% Mismatches: 0
Query Match: 97.57% Indels: 38
DB: 9 Gaps: 1

US-09-954-846-2 (1-258) x US-09-905-291A-206 (1-1620)

QY 1 MetAlaValLeuAlaProLeuIleAlaLeuValTySerValProArgLeuSerArgTrp 20
DB 3 ATGGCGGTCTGGCACCTCTAAATGCTCTCGTGTATTGGTGGCGGCACTTTCACGATGG 62
QY 21 LeuAlaGlnProTyrTyrLeuLeuSerAlaLeuLeuSerAlaAlaPheLeuValArg 40
DB 63 CTGCCCCAACCTTACTACTACCTTCTGCGGCCCTGCTCTCTGCGGCTTCTTACTCTGAGG 122
QY 41 LysLeuProLeuCysHisGlyLeuProThrGlnArgGluAspGlyAsnProCysasp 60
DB 123 AAACCTGCGCGGCTCTGCCACGGTCTGCCACCCACCGAAGACCGTAACCCGCTGTGAC 182
QY 61 PheAspTrpArgGluValGluIleLeuMetPheLeuSerAlaIleValMetMetLysAsn 80
DB 183 TTTGACTGGAGAAAGTGGAGATCCTGATGTTTCTCAGTCCCATTTGATGATGATGAAGAC 242
QY 81 ArgArgSer----- 83
DB 243 CGCAGATCCATCACTGTGGAGACACATATAGGCAACATTTTCATGTTTAGTAAGATGCC 302
QY 83 ----- 83
DB 303 AACACAATTCTTTCTTCGCTTGGATATTTCGCATGGCGCCTACTTTACATCACACTGCG 362

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT19
CLONE: 3244141
SEQUENCE DESCRIPTION: SEQ ID NO: 4 :
US-09-954-846-4

Query Match
Best Local Similarity 100.0%; Score 1555; DB 10; Length 1555;
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGAGGCGGGGAGACCTACGACCGCGGAGCAGTGGCCGTACGGCGCGAAAGAT 60
DB 1 AGCGAGGCGGGGAGACCTACGACCGCGGAGCAGTGGCCGTACGGCGCGAAAGAT 60
61 GCGGCTTTGGGCACTTAATTGCTCTCGTGTATTCGGTGGCGGACTTTACGATGGCT 120
DB 61 GCGGCTTTGGGCACTTAATTGCTCTCGTGTATTCGGTGGCGGACTTTACGATGGCT 120
121 CGCCCAACCTTACTACCTTCTGCGGCCCTGCTCTCTGCTGCTTCTACTCGTGAGAA 180
DB 121 CGCCCAACCTTACTACCTTCTGCGGCCCTGCTCTCTGCTGCTTCTACTCGTGAGAA 180
181 ACTGCGCGCGCTCTGCGGCCCTGCTCTCTGCTGCTTCTACTCGTGAGAA 240
DB 181 ACTGCGCGCGCTCTGCGGCCCTGCTCTCTGCTGCTTCTACTCGTGAGAA 240
241 TGACTGGAGAGAGTGGAATCTGATGTTTCTCAGTGCCATTTGATGATGAAGAACCG 300
DB 241 TGACTGGAGAGAGTGGAATCTGATGTTTCTCAGTGCCATTTGATGATGAAGAACCG 300
301 CAGATCCATCTTCTGATGAGTGAACCAACCCCTTATATATGCGCCCTGAGTATCAA 360
DB 301 CAGATCCATCTTCTGATGAGTGAACCAACCCCTTATATATGCGCCCTGAGTATCAA 360
361 GTACTTCAATGATAAACCATTGATGAGTGAACCAACCCCTTATATATGCGCCCTGAGTATCAA 420
DB 361 GTACTTCAATGATAAACCATTGATGAGTGAACCAACCCCTTATATATGCGCCCTGAGTATCAA 420
421 TGTGGAGTCTTTGGCAATTTGCTTAATGATGCGCAATCATTTGCCCTATCTATCTGTA 480
DB 421 TGTGGAGTCTTTGGCAATTTGCTTAATGATGCGCAATCATTTGCCCTATCTATCTGTA 480
481 CCTCTCCCTTAATACAACTGACAGGCTAAATTTGGGAAGTGTGTCGACCTA 540
DB 481 CCTCTCCCTTAATACAACTGACAGGCTAAATTTGGGAAGTGTGTCGACCTA 540
541 TACTGATGTTAGTACGCGGTACAAAGTGAGCAGATCACCTACCAAGCAACTCCCTAC 600
DB 541 TACTGATGTTAGTACGCGGTACAAAGTGAGCAGATCACCTACCAAGCAACTCCCTAC 600
601 CTTGATCCTGTTCAAGGTGGCAAGGAGGAAATGCGCGGCGCAAGATTGACAAAGG 660
DB 601 CTTGATCCTGTTCAAGGTGGCAAGGAGGAAATGCGCGGCGCAAGATTGACAAAGG 660
661 ACGGCTGCTCATGAGCTTCTCTGAGGAGATGATCGGAGATTTACITTAATGA 720
DB 661 ACGGCTGCTCATGAGCTTCTCTGAGGAGATGATCGGAGATTTACITTAATGA 720
721 GCTATACCAAGCGGCGCAAGAACTATCAAGGCTGGAGCAATATCCCTGAGGAGCACC 780
DB 721 GCTATACCAAGCGGCGCAAGAACTATCAAGGCTGGAGCAATATCCCTGAGGAGCACC 780
781 TGTGGCTTCAACCCCAACCACTGATGAGTGGGAAACCAAGAGGATAAATGAATCC 840
DB 781 TGTGGCTTCAACCCCAACCACTGATGAGTGGGAAACCAAGAGGATAAATGAATCC 840
841 TCACCTTTGGCAGTGTCTCTCTGCAATTCAGGCTTTTCCATACCAAGCCTG 900
DB 841 TCACCTTTGGCAGTGTCTCTCTGCAATTCAGGCTTTTCCATACCAAGCCTG 900

QY 901 AGGCTGAGCGCTTTTATTTATGTTTCCCTTGGCTGTGACTGGGTGGGCGAGCATGCG 960
DB 901 AGGCTGAGCGCTTTTATTTATGTTTCCCTTGGCTGTGACTGGGTGGGCGAGCATGCG 960
961 CTTCTGATTTTAAAGAGGCGATCTAGGGAATTTGTGAGCAGCCCTCAGAGAAAGCCTGCCAT 1020
DB 961 CTTCTGATTTTAAAGAGGCGATCTAGGGAATTTGTGAGCAGCCCTCAGAGAAAGCCTGCCAT 1020
1021 GCTGTGCGCAACTGTTTCACTGGAGCAAGAGATCTCATAGGACGGGGAATG 1080
DB 1021 GCTGTGCGCAACTGTTTCACTGGAGCAAGAGATCTCATAGGACGGGGAATG 1080
1081 GTTTCCTTCCCAAGCTTGGGTGAGTGTGTAAGTCTTATCAGCTATTCAGACATCTCCAT 1140
DB 1081 GTTTCCTTCCCAAGCTTGGGTGAGTGTGTAAGTCTTATCAGCTATTCAGACATCTCCAT 1140
1141 GGTTCCTTCCCAAGCTTGGGTGAGTGTGTAAGTCTTATCAGCTATTCAGACATCTCCAT 1200
DB 1141 GGTTCCTTCCCAAGCTTGGGTGAGTGTGTAAGTCTTATCAGCTATTCAGACATCTCCAT 1200
1201 TTAGACCTAGATTTAACCTTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCT 1260
DB 1201 TTAGACCTAGATTTAACCTTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCTAAGCT 1260
1261 CCAAGGACTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1320
DB 1261 CCAAGGACTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1320
1321 AAGCTTAACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 AAGCTTAACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1381 AAGCAATATTTGAACCAACCCCTTATTTGTTGGGATTTGAGAGGGGTGAATAGAGGCTTG 1440
DB 1381 AAGCAATATTTGAACCAACCCCTTATTTGTTGGGATTTGAGAGGGGTGAATAGAGGCTTG 1440
1441 AGACTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 AGACTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
1501 TGACATCT 1555
DB 1501 TGACATCT 1555

RESULT 2

US-09-925-300-751
; Sequence 751, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 751
; LENGTH: 1598
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-300-751

Query Match 97.2%; Score 1511; DB 10; Length 1598;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1511; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 42 CCGTACGCGCGCAAGAGATGCGGCTTTCGACCTCTAATGCTCTCTGCTATTCGGTGC 101

Db 26 CGGTACGGCCGAAAGATGGGGTCTTGGACCTCTAATGCTCTCGGTATATCGGTGC 85
Qy 102 CGGACTTTCAGATGGCTCGCCCAACCTTACTACTCTGTGCGGCCCTGCTCTCTGCTG 161
Db 86 CGGACTTTCAGATGGCTCGCCCAACCTTACTACTCTGTGCGGCCCTGCTCTCTGCTG 145
Qy 162 CTTCTCTACTCGTGAGGAACATGCGCGCTCTGCCACGGTCTGCCACCCCAACGCGAAG 221
Db 146 CTTCTCTACTCGTGAGGAACATGCGCGCTCTGCCACGGTCTGCCACCCCAACGCGAAG 205
Qy 222 ACCTTAACCCCTGTGACTTGTACTGGAGAGAGTGGAGATCTGATGTTCTCAGTGCCA 281
Db 206 ACCTTAACCCCTGTGACTTGTACTGGAGAGAGTGGAGATCTGATGTTCTCAGTGCCA 265
Qy 282 TTGTGATGATGAAGAACCGCAGATCCATGTTCTGATGACGTGCAAAACCCCTCTATATA 341
Db 266 TTGTGATGATGAAGAACCGCAGATCCATGTTCTGATGACGTGCAAAACCCCTCTATATA 325
Qy 342 TGGGCCCTGAGTATATCAAGTACTTCAATGATPAAACCATTTGATGAGGAACCTAGAACGG 401
Db 326 TGGGCCCTGAGTATATCAAGTACTTCAATGATPAAACCATTTGATGAGGAACCTAGAACGG 385
Qy 402 ACAAGAGGCTCAGTTGGATTTGGAGTCTTTTGCCAATTTGGTCTTAATGACTGCCAATCAT 461
Db 386 ACAAGAGGCTCAGTTGGATTTGGAGTCTTTTGCCAATTTGGTCTTAATGACTGCCAATCAT 445
Qy 462 TTGCCCTTATCTATGCTGACCTCTCCCTTAATACAACTGACAGGGCTAAATTTGGGA 521
Db 446 TTGCCCTTATCTATGCTGACCTCTCCCTTAATACAACTGACAGGGCTAAATTTGGGA 505
Qy 522 AGGTGGATGTGGACGCTATCTATGTTAGTACGGGTACAAAGTGAGCACATCACCC 581
Db 506 AGGTGGATGTGGACGCTATCTATGTTAGTACGGGTACAAAGTGAGCACATCACCC 565
Qy 582 TCACCAAGCAACTCCCTACCTGTATCCAGTTCCTTCCAAAGTGGCAAGGAGCAATGCGCGGC 641
Db 566 TCACCAAGCAACTCCCTACCTGTATCCAGTTCCTTCCAAAGTGGCAAGGAGCAATGCGCGGC 625
Qy 642 CACAGATGACAAGAAAGAGGGCTGCTCATGACCTTCTCTGAGGAGATGTATCC 701
Db 626 CACAGATGACAAGAAAGAGGGCTGCTCATGACCTTCTCTGAGGAGATGTATCC 685
Qy 702 GAGATTTAACTTAATGAGCTATACACGCGGCCCAAGAACTATCAAGGCTGGAGACA 761
Db 686 GAGATTTAACTTAATGAGCTATACACGCGGCCCAAGAACTATCAAGGCTGGAGACA 745
Qy 762 ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACCCACACAGTGTAGATGGGAAACA 821
Db 746 ATATCCCTGAGGAGCAGCTGTGGCTTCAACCCCAACCCACACAGTGTAGATGGGAAACA 805
Qy 822 AGAAGGATAAATAGATCTCCTTGGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 881
Db 806 AGAAGGATAAATAGATCTCCTTGGCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 865
Qy 882 TTCCATACCAACAGCCTGAGGCTGCAGCTTTTATTTATGTTTCCCTTTGGCTGTGAC 941
Db 866 TTCCATACCAACAGCCTGAGGCTGCAGCTTTTATTTATGTTTCCCTTTGGCTGTGAC 925
Qy 942 TGGGTGGGGCAGCATGCACTCTGATTTTAAAGAGGCATCTAGGAAATTTGTCAGGCACC 1001
Db 926 TGGGTGGGGCAGCATGCACTCTGATTTTAAAGAGGCATCTAGGAAATTTGTCAGGCACC 985
Qy 1002 CTACAGGAGCCTGCGCATGCTGTGGCCCAACTGTTTCACTGGAGCAAGAAAGAGATCTCA 1061
Db 986 CTACAGGAGCCTGCGCATGCTGTGGCCCAACTGTTTCACTGGAGCAAGAAAGAGATCTCA 1045
Qy 1062 TAGGACGAGGGGAAATGGTTTCCCTCAAGCTTGGTCTAGTGTCTAACTGCTTATCA 1121
Db 1046 TAGGACGAGGGGAAATGGTTTCCCTCAAGCTTGGTGTAGTGTCTAACTGCTTATCA 1105
Qy 1122 GCTATTGAGACATCTCCATGCTTTTCCATGAACTCTGTGTTTCAATCATCTCTCTCTTA 1181

Db 1106 GCTATTGACAGATCTCCATGGTTTCTCCATGAACACTCTCTGTGGTTTCATCATCTCTCTTA 1165
Qy 1182 GTTGACCTGCACAGCTTGTGTAGACCTAGATTTAACCTTAAGCTAAGATGCTGGGTATA 1241
Db 1166 GTTGACCTGCACAGCTTGTGTAGACCTAGATTTAACCTTAAGCTAAGATGCTGGGTATA 1225
Qy 1242 GAACGCTAAGAAATTTTCCCCCAAGGACTCTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTA 1301
Db 1226 GAACGCTAAGAAATTTTCCCCCAAGGACTCTGCTTCCCTTAAGCCCTTCTGGCTTCGTTTA 1285
Qy 1302 TGGTCTTCAATTAAGAGTAAAGCCTAACCTTTCGCTAGTCTTCCCTTAAGGAGAAACCTTTAAC 1361
Db 1286 TGGTCTTCAATTAAGAGTAAAGCCTAACCTTTCGCTAGTCTTCCCTTAAGGAGAAACCTTTAAC 1345
Qy 1362 CACAAGATTTTATCATTTGAAGACATATTTGAACAACCCCTATTTTGTGGGATTCGAGA 1421
Db 1346 CACAAGATTTTATCATTTGAAGACATATTTGAACAACCCCTATTTTGTGGGATTCGAGA 1405
Qy 1422 AGGGTGAATAGAGGCTTGAGACTTTTCCCTTGTGTGTAGGACTTGGAGGAGAAATCCCC 1481
Db 1406 AGGGTGAATAGAGGCTTGAGACTTTTCCCTTGTGTGTAGGACTTGGAGGAGAAATCCCC 1465
Qy 1482 TGGACTTTTCACTAACCTCTGACATACTCCCAACCCAGTGTGATGGCTTCCGTAATAA 1541
Db 1466 TGGACTTTTCACTAACCTCTGACATACTCCCAACCCAGTGTGATGGCTTCCGTAATAA 1525
Qy 1542 AAAGATTTGGGATT 1554
Db 1526 AAAGATTTGGGATT 1538

RESULT 3
US-09-852-797-19
Sequence 19, Application US/09852797
Patent No. US20020172994A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P2
CURRENT APPLICATION NUMBER: US/09/852,797
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 60/265,583
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/152,060
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: PCT/US98/04858
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/040,762
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/040,710
PRIOR FILING DATE: 1997-03-14
PRIOR APPLICATION NUMBER: 60/050,934
PRIOR FILING DATE: 1997-03-30
PRIOR APPLICATION NUMBER: 60/048,100
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,357
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/048,189
PRIOR FILING DATE: 1997-05-30
PRIOR APPLICATION NUMBER: 60/057,765
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/048,970
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 1699
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE

